BASIC

MOLECULAR GENETICS

Dr. Rokaya Shalaby

Assistant Professor

Of

Cytogenetics and Molecular Biology

2006

CONTENTS

Introduction

The great discoveries cataloged in our history have mostly been achieved by exploring our environment, whether it be Eratosthenes calculating the Earth's circumference, Columbus sailing across the Atlantic, Galileo peering at the moons of Jupiter, or Darwin studying finches in the Galapagos. After many centuries, we have built up an approximate understanding of our external universe, but the universe within us has only very recently been the subject of serious study. The application of microscopy to the study of cells and sub-cellular structures provided one major route into this world, to be followed by pioneering advances in biochemistry and then molecular biology. Now, as we enter the next millennium, we are on the threshold of a truly momentous achievement that will have enormous implications for the future. For the first time, we will know our genetic endowment the sequence of our DNA. Then our voyage into the universe within really will have begun. Sequencing our DNA will be just the beginning of a huge effort to understand exactly how this sequence can specify a person, and how the DNA of other organisms is related to us and to their biologies.

This book is a brief introduction for molecular genetic. It is designated for both under and post-graduated students. As further aid, more than 70 diagrams and photographs were supplemented to provide the key for understanding the text.

HISTORY of GENETICS

Year	Discovery
1859	*Discovery: Natural Selection.*
	Charles Darwin wrote "On the Origin of Species by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life."

Year	Discovery
1865	*Discovery: Heredity Transmitted in Units*
	Gregor Mendel's experiments on peas demonstrate that heredity is transmitted in discrete units. The understanding that genes remain distinct entities even if the characteristics of parents appear to blend in their children explains how natural selection could work and provides support for Darwin's proposal.

Year	Discovery
1869	*Discovery: DNA Isolated.*
	Frederick Miescher isolates DNA from cells for the first time and calls it "nuclein".

Year	Discovery
1879	*Discovery: Mitosis Described*
	Walter Flemming describes chromosome behavior during animal cell division. He stains chromosomes to observe them clearly and describes the whole process of mitosis in 1882.

Year	Discovery
1900	*Discovery: Rediscovery of Mendel's work.*
	Botanists DeVries, Correns, and von Tschermak independently rediscover Mendel's work while doing their own work on the laws of inheritance. The increased understanding of cells and chromosomes at this time allowed the placement of Mendel's abstract ideas into a physical context.

1902	*Discovery: Chromosome Theory of Inheritance.*
	Walter Sutton observes that the segregation of chromosomes during meiosis matched the segregation pattern of Mendel's

1902	*Discovery: Orderly Inheritance of Disease*
	A British physician, Archibald Garrod, observes that the disease alkaptonuria is inherited according to Mendelian rules. This disease involves a recessive mutation, and was among the first conditions ascribed to a genetic cause.

1909	*Discovery: The Word Gene is Coined.*
	Wilhelm Johannsen coins the word "gene" to describe the Mendelian unit of heredity. He also uses the terms genotype and phenotype to differentiate between the genetic traits of an individual and its outward appearance.

1911	*Discovery: Chromosomes Carry Genes.*
	Thomas Hunt Morgan and his students study fruit fly chromosomes. They show that chromosomes carry genes, and also discover genetic linkage.

1941	*Discovery: One Gene, One Enzyme Hypothesis.*
	George Beadle and Edward Tatum's experiments on the red bread mold, *Neurospora crassa,* show that genes act by regulating distinct chemical events. They propose that each gene directs the formation of one enzyme.

1943	*Discovery: DNA Has a Regular Periodic Structure.*
	William Astbury, a British scientist, obtains the first X-ray diffraction pattern of DNA, which reveals that DNA must have a regular periodic structure. He suggests that nucleotide bases are stacked on top of each other.

1944	*Discovery: DNA Transforms Cells.*
	Oswald Avery, Colin MacLeod, and Maclyn McCarty show that DNA (not proteins) can transform the properties of cells, thus clarifying the chemical nature of genes.

1944	*Discovery: Jumping Genes.*
	Barbara McClintock, using corn as the model organism, discovers that genes can move around on chromosomes. This shows that the genome is more dynamic than previously thought. These mobile gene units are called transposons and are found in many species.

1952	*Discovery: Genes are Made of DNA.*
	Alfred Hershey & Martha Chase show that only the DNA of a virus needs to enter a bacterium to infect it, providing strong support for the idea that genes are made of DNA.

1953	*Discovery: DNA Double Helix.*
	Francis H. Crick and James D. Watson described the double helix structure of DNA. They receive the Nobel Prize for their work in 1962 ·

1955	*Discovery: 46 Human Chromosomes.*
	Joe Hin Tjio defines 46 as the exact number of chromosomes in human cells.

1955	*Discovery: DNA Copying Enzyme.*
	Arthur Kornberg and colleagues isolated DNA polymerase, an enzyme that would later be used for DNA sequencing.

1956	*Discovery: Cause of Disease Traced to Alteration*
	Vernon Ingram discovers that a specific chemical alteration in a hemoglobin protein is the cause of sickle cell disease.

1958	*Discovery: Semiconservative Replication of DNA.*
	Matthew Meselson and Franklin Stahl demonstrate that DNA replicates semiconservatively: each strand from the parent DNA molecule ends up paired with a new strand from the daughter generation.

1959	*Discovery: Chromosome Abnormalities Identified.*
	Jerome Lejeune and his colleagues discover that Down Syndrome is caused by trisomy 21. There are three copies, rather than two, of chromosome 21, and this extra chromosomal material interferes with normal development.

1961	*Discovery: First Screen for Metabolic Defect in Newborns.*
	Robert Guthrie develops a method to test newborns for the metabolic defect, phenylketonuria (PKU).

1961	*Discovery: mRNA Ferries Information.*
	Sydney Brenner, François Jacob and Matthew Meselson discover that mRNA takes information from DNA in the nucleus to the protein-making machinery in the cytoplasm.

1966	*Discovery: Genetic Code Cracked*
	Marshall Nirenberg and others figure out the genetic code that allows nucleic acids with their 4 letter alphabet to determine the order of 20 kinds of amino acids in proteins.

1968	*Discovery: First Restriction Enzyme Described.*
	Scientists describe restriction nucleases, enzymes that recognize and cut specific short sequences of DNA. The resulting fragments can be used to analyze DNA, and these enzymes later became an important tool for mapping genomes.

1972	*Discovery: First Recombinant DNA.*
	Scientists produce recombinant DNA molecules by joining DNA from different species and subsequently inserting the hybrid DNA into a host cell, often a bacterium.

1973	*Discovery: First Animal Gene Cloned.*
	Researchers fuse a segment of DNA containing a gene from the African clawed frog Xenopus with DNA from the bacterium E. coli and placed the resulting DNA back into an E. coli cell. There, the frog DNA was copied and the gene it contained directed the production of a specific frog protein.

1975	*Discovery: DNA Sequencing.*
	Two groups, Frederick Sanger and colleagues, and Alan Maxam and Walter Gilbert, both develop rapid DNA sequencing methods. The Sanger method is most commonly employed in the lab today, with colored dyes used to identify each of the four nucleic acids that make up DNA.

1976	*Discovery: First Genetic Engineering Company.*
	Herbert Boyer founds Genentech. The company produces the first human proteinin a bacterium, and by 1982 markets the first recombinant DNA drug, human insulin.

1977	*Discovery: Introns Discovered.*
	Richard Roberts' and Phil Sharp's labs show that eukaryotic genes contain many interruptions called introns. These noncoding regions do not directly specify the amino acids that make protein products.

1981	*Discovery: First Transgenic Mice and Fruit Flies.*
	Scientists successfully add stably inherited genes to laboratory animals. The resulting transgenic animals provide a new way to test the functions of genes.

1982	*Discovery: Gene Bank Database Formed.*
	Scientists begin submitting DNA sequence data to a National Institutes of Health (NIH) database that is open to the public.

1983	*Discovery: First Disease Gene Mapped*
	A genetic marker for Huntington's disease is found on chromosome 4.

1983	*Discovery: PCR Invented.*
	The polymerase chain reaction, or PCR, is used to amplify DNA. This method allows researchers to quickly make billions of copies of a specific segment of DNA, enabling them to study it more easily.

1986	*Discovery: First Time a Disease Gene is Positionally Cloned.*
	A method for finding a gene without the knowledge of the protein it encodes is developed. So called, positional cloning can help in understanding inherited disease, such as muscular dystrophy.

1987	*Discovery: First Human Genetic Map*
	The first comprehensive genetic map is based on variations in DNA sequence that can be observed by digesting DNA with restriction enzymes. Such a map can be used to help locate genes responsible for diseases.

1987	*Discovery: Yeast Artificial Chromosomes.*
	Scientists discover that artificial chromosomes made from yeast can reliably carry large fragments of human DNA containing millions of base-pair pieces. Earlier methods used plasmids and viruses, which can carry only a few thousand base-pair pieces. The ability to deal with much larger pieces of DNA makes mapping the human genome easier.

1989	*Discovery: Microsatellites Are New Genetic Markers*
	Repetitive DNA sequences called microsatellites are used as genetic landmarks to distinguish between people. Another type of marker, sequence–tagged sites, are unique stretches of DNA that can be used to make physical maps of human chromosomes.

1990	*Discovery: Launch of the Human Genome Project.*
	The Department of Energy and the National Institutes of Health announce a plan for a 15-year project to sequence the human genome. This will eventually result in sequencing all 3.2 billion letters of the human genome.

1991	*Discovery: ESTs, Fragments of Genes.*
	An expressed-sequence tag (EST) an identified piece of a gene, is made by copying a portion of a messenger RNA (mRNA) molecule. As such, ESTs provide a way to focus on the "expressed" portion of the genome, which is less than one-tenth.

1992	*Discovery: Second-Generation Genetic Map of Human Genome*
	A French team builds a low-resolution, microsatellite genetic map of the entire human genome. Each generation of the map helps geneticists more quickly locate disease genes on chromosomes.

1994	*Discovery: FLAVR SAVR Tomato.*
	The Food And Drug Administration approves the sale of the first genetically modified food.

1995	Discovery: Ban on Genetic Discrimination in the Workplace
	Protection under the American with Disabilities Act is extended to cover discrimination based on genetic information.

1996	*Discovery: Mouse Genetic Map Completed.*
	The lab mouse is valuable for genetics research because humans and mice share almost all of their genes, and the genes on average are 85% identical. The mouse genetic map increases the utility of mice as animal models for genetic disease in humans.

1997	*Discovery: E. coli Genome Sequenced.*
	The complete sequence of the *E. coli* genome will help scientists learn even more about this extensively studied bacterium ·

1998	*Discovery: M. tuberculosis Bacterium Sequenced*
	Mycobacterium tuberculosis causes the chronic infectious disease tuberculosis. The sequencing of this bacterium is expected to help scientists develop new therapies to treat the disease.

1998	*Discovery: Roundworm C. elegans Sequenced*
	The first genome sequence of a multicellular organism, the round worm, Caenorhabditis elegans, is completed.

1999	*Discovery: Chromosome 22 Sequenced*
	The first finished, full-length sequence of a human chromosome is produced. Chromosome 22 was chosen to be first because it is relatively small and had a highly detailed map already available. Such a map is necessary for the clone by clone sequencing approach.

2000	*Discovery: Human Genome Working Draft Completed*
	By the end of spring 2000, Human Genome Project (HGP) researchers sequence 90 percent of the human genome with 4-fold redundancy. This working draft sequence is estimated to be 99.9% accurate.

2002	*Discovery: Mouse Genome Working Draft Assembled and Analyzed*
	The Mouse Genome Sequencing Consortium publishes an assembled draft and comparative analysis of the mouse genome. This milestone was originally planned for 2003.

2002	*Discovery: Rat Genome Working Draft Completed*
	By Fall 2002, researchers sequence over 90% of the rat genome with over 5-fold redundancy.

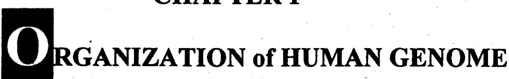
2003	*Discovery: Completion of the Human Genome Sequencing*
	The finished human genome sequence will be at least 99.99% accurate.

2004	*Discovery:* Trim Count of Human Genes to 20,000-25,000
	The International Human Genome Sequencing Consortium, led in the United States by the National Human Genome Research Institute (NHGRI) and the Department of Energy (DOE), today published its scientific description of the finished human genome sequence, reducing the estimated number of human protein-coding genes from 35,000 to only 20,000-25,000, a surprisingly low number for our species.

The Future

It will take decades of research for scientists to understand all of the information that is contained within the human genome. In time, more human diseases will be understood at the level of the molecules that are involved, which could dramatically change the practice of medicine by leading to the development of new drugs, as well as to genetic testing to improve and individualize treatments.

CHAPTER 1



ORGANIZATION of HUMAN GENOME

Introduction

The human genome is the term used to describe the total genetic information (DNA content) in human cells. It really comprises two genomes: a complex nuclear genome which accounts for 99.9995% of the total genetic information, and a simple mitochondrial genome which accounts for the remaining 0.0005%. The nuclear genome provides the great bulk of essential genetic information, most of which specifies polypeptide synthesis on cytoplasmic ribosomes. Mitochondria possess their own ribosomes and the few polypeptide-encoding genes in the mitochondrial genome produce mRNAs which are translated on the mitochondrial ribosomes.

Pseudogenes,

They are truncated gene copies and gene fragments which are commonly found in multigene families. Families of RNA genes or polypeptide-encoding genes are frequently characterized by defective copies of essentially all of the gene or its coding sequence (pseudogenes), or a portion of it, in some cases a single exon (gene fragments). A large variety of different classes are found **(Fig.1)**.

Gene superfamilies

They are genes which encode products (proteins) that are functionally related. They show only very weak sequence homology over a large segment, and, there are some evidence for

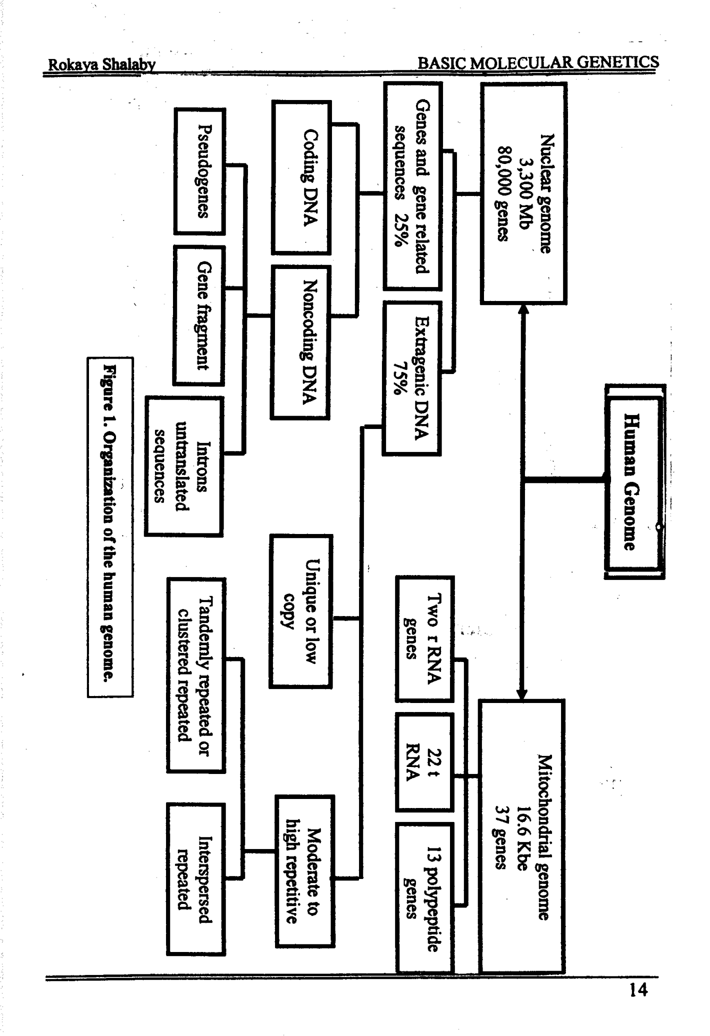


Figure 1. Organization of the human genome.

general common structural features. For example, in the immune system there are immunoglobulin gene family which responsible for immunoglobuline protein production , other related genes such as the HLA genes, TCR genes (share in antigen recognition), genes are known to encode proteins with a domain structure that resembles that of immunoglobulins and their function related to the immune system. Although, therefore, the level of sequence homology between such genes may be very low, the similarities in function and general domain structure have suggested the existence of a so-called Ig superfamily.

Human genome contains a large amount of highly repeated DNA sequence families, which are largely transcriptionally inactive. Noncoding repetitive DNA shows two major types of organization: tandemly repeated and interspersed.

Tandemly repeated noncoding DNA

Such families are defined by **blocks** (or arrays) of tandemly repeated DNA sequences. Individual arrays can occur at a few or many different chromosomal locations. Depending on the average size of the arrays of repeat units, highly repetitive noncoding DNA can be grouped into three subclasses: **satellite, minisatellite** and **microsatellite DNA.**

Satellite DNA (10 -15 % of repetitive DNA)

It is composed of **very long** arrays of tandem repeats which can be separated from bulk DNA by buoyant density gradient centrifugation. Following centrifugation, they appear as minor (satellite) bands of different buoyant density from a major band which represents bulk DNA. Human satellite DNA is comprised of very large arrays of tandemly repeated DNA with the repeat unit being a simple or moderately complex sequence. Repeated DNA of this type is not transcribed and accounts for the bulk of

the heterochromatic regions of the genome, being notably found in the vicinity of the centromeres (pericentromeric heterochromatin).

Minisatellite

DNA comprises a collection of **moderately** sized arrays of tandemly repeated DNA sequences which are **dispersed** over considerable portions of the nuclear genome. Like **satellite** DNA sequences, they are not normally transcribed.

Hypervariable minisatellite DNA

Hypervariable minisatellite DNA **sequences** are **highly** polymorphic and are organized in over **1000** arrays (0.1–20 kb long) of short tandem repeats. The **repeat** units in **different** hypervariable arrays vary considerably in size, but share a common core sequence, GGGCAGGAXG (where X = any nucleotide), which is similar in size and in G content. While many of the arrays are found near the telomeres, several hypervariable minisatellite DNA sequences occur at other chromosomal locations **(Fig.2)**. The great majority of hypervariable minisatellite DNA sequences are not transcribed, except for elements occuring within noncoding intragenic sequences. Some, however, are expressed.

The significance of hypervariable minisatellite DNA is not clear, although it has been reported to be a 'hotspot' for homologous recombination in human cells. Nevertheless it has found many applications. Various individual loci have been characterized and used as **genetic markers**, although the preferential localization in subtelomeric regions has limited their use for genome-wide linkage studies.

DNA fingerprinting

A major application has been in DNA fingerprinting, in which a single DNA probe which contains the common core sequence can hybridize simultaneously to multiple minisatellite DNA loci on all chromosomes, resulting in a complex individual-specific hybridization pattern.

Telomeric DNA

Another major family of minisatellite DNA sequences is found at the termini of chromosomes. **(Fig.2)** The principal constituent of telomeric DNA is 10–15 kb of tandem hexanucleotide repeat units, especially TTAGGG, which are added by telomerase enzyme.

By acting as buffers to protect the ends of the chromosomes from degradation and loss and by providing a mechanism for replicating the ends of the linear DNA of chromosomes, these simple repeats are directly responsible for telomere function.

Microsatellite DNA:

DNA families include small arrays of tandem repeats which are simple sequence (1–4 bp) and are interspersed throughout the genome **(Fig.2)**.

Of the mononucleotide repeats, runs of A and of T are very common and together account for about 0.3% of the nuclear genome. By contrast, runs of G and of C are very rare.

In the case of dinucleotide repeats, arrays of CA repeats (TG repeats on the complementary strand) are very common, accounting for 0.5% of the genome, and are often highly polymorphic.

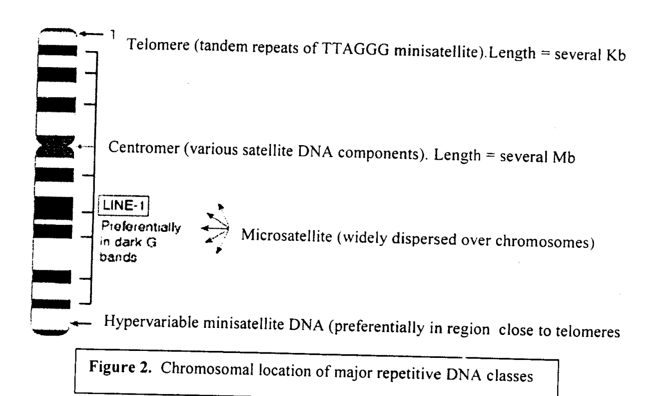


Figure 2. Chromosomal location of major repetitive DNA classes

The significance of microsatellite DNA is not known. Although microsatellite DNA has generally been identified in intergenic DNA or within the introns of genes, a few examples have been recorded within the coding sequences of genes.

Tandem repeats of three nucleotides in coding DNA may be sites that are prone to pathogenic expansions.

In addition to these three major (satellite, minisatellite and microsatellite) a fourth class has recently been recognized and described as **megasatellite** or macrosatellite DNA: Despite the name, this type of DNA is characterized by array lengths which can be comparatively modest compared to some satellite DNA arrays. Instead, the prefix mega- has been used to emphasize the large size of the repeating unit which can be several kilobases long.

Interspersed repetitive non-coding DNA

The individual repeat units are not clustered, but are dispersed at numerous locations in the genome, and together account for perhaps one third of the DNA in the human genome. Two major classes of mammalian interspersed repetitive DNA families have been discerned on the basis of repeat unit length :SINEs and LINEs.

SINEs (Short Interspersed Nuclear Elements)

The most conspicuous human SINE is the Alu repeat family (so called because of early attempts at characterizing the sequence using the restriction nuclease Alu I). The Alu repeat contains an internal RNA polymerase III promoter sequence. It has attained a very high copy number in the human genome. The Alu repeat is primate-specific but other mammals have similar types of sequence **(Fig.3)**.

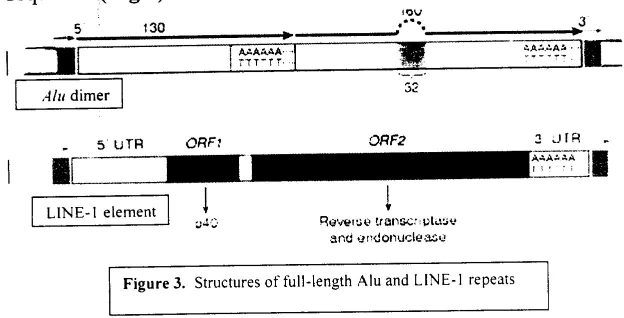


Figure 3. Structures of full-length Alu and LINE-1 repeats

LINEs (Long Interspersed Nuclear Elements)

Human LINEs are exemplified by the LINE-1 or L1 element (also called the Kpn repeat because of early attempts at

characterizing this family using the restriction nuclease Kpn I. The LINE-1 element is also found in other mammals such as the mouse **(Fig.3)**.

In addition to the human Alu and LINE-1 repeat families, there are many smaller families, including the THE-1 (transposable human element family), many MER (medium reiteration frequency) families and families of human endogenous retroviruses (HERV) or retrovirus-like elements (RTLV).

Distribution of human genes

The nuclear genome contains about 65,000–100,000 genes but only about 3% of the genome represents coding sequences. The number of genes in the human genome has been the subject of much speculation; while the small mitochondrial genome is known to have precisely 37 genes. The number of these genes have been suggested on the basis of three approaches:

1. **Genomic sequencing**: The sequencing of large chromosomal regions may suggest that there are about 70,000 genes. This is based on the observation that gene-rich regions have an average gene density of close to one per 20 kb, but gene-poor regions have a much lower density, (one-tenth of this density), and that the genome is split 50:50 into gene-rich and gene-poor regions.
2. **CpG island number:** Restriction enzyme analysis using the methylation-sensitive enzyme HpalI suggests that the total number of CpG islands in the human genome is 45, 000.
3. **EST analysis**: Large-scale random sequencing of cDNA clones provides so-called expressed sequence tags. Comparison of known human EST sequences with a large set of different

human genomic coding DNA sequences listed in sequence databases has suggested a figure of about 65,000 human genes

RNA-encoding gene families

While the great majority of human genes are expected to encode polypeptides, a significant minority encode mature RNA molecules of diverse function.

Ribosomal RNA (rRNA) genes

There are multiple rRNA genes. Cytoplasmic rRNAs are encoded by a single transcription unit , which is tandemly repeated about 250 times, comprising five clusters of about 50 tandem repeats located on the short arms of human chromosomes 13, 14, 15, 21 and 22. In addition, the 5S cytoplasmic rRNA is encoded by several hundred gene copies in at least three clusters on the long arm of chromosome 1. The major rationale for the repetition of cytoplasmic rRNA genes is likely to be based on gene dosage: by having a comparatively large number of these genes, the cell can satisfy the huge demand for cytoplasmic ribosomes needed for protein synthesis.

Small nuclear RNA (snRNA) genes

A large dispersed family of genes encodes a heterogeneous collection of several hundred small nuclear RNA species.

Transfer RNA (tRNA) genes

These belong to a very large dispersed gene family, comprising more than 40 different subfamilies each with several members which encode the different species of cytoplasmic tRNA. In addition to multiple copies of genes specifying the individual cytoplasmic tRNA molecules, there are several defective gene copies (pseudogenes).

Human multigene families

DNA sequences in the nuclear diploid genome usually exist as two allelic copies (on paternal and maternal homologous chromosomes). In addition to this degree of repetition, about 40% of the human nuclear genome in both haploid and diploid cells is composed of sets of closely related nonallelic DNA sequences (DNA sequence families or repetitive DNA).

The human genome consists of three broad sequence components:

1. **Single copy**, or at least very low copy number, DNA (60%) reassociates very slowly. A single strand from a single copy sequence will require some considerable time to find a complementary partner strand, given that the vast majority of DNA fragments are unrelated to it.
2. **Moderately repetitive** (30%) - intermediate speed of reassociation.
3. **Highly repetitive** (10%) - reassociates very rapidly. There are numerous copies of the same sequence and the chances of quickly finding complementary partners within the mass of different fragments are high.

Functionally similar genes

As seen in the previous section, some families of RNA genes are clustered. In the case of polypeptide-encoding gene families, some genes encoding identical or functionally related products are clustered, but often they are dispersed on several chromosomes.

Functionally identical genes

A very few human polypeptides are known to be encoded by two or more identical gene copies. Often, these are encoded by recently duplicated genes in a gene cluster, as in the case of the duplicated alpha-globin genes. In addition, some genes on different chromosomes encode identical polypeptides e.g. histone gene subfamilies. Histones can be classified into five groups in terms of structure: H1 (the linker histone) and the four core histones, H2A, H2B, H3 and H4. In addition histone genes can be classified into three groups according to expression: (i) replication-dependent (restricted to the S phase of the cell cycle); (ii) replication-independent (expressed at a low level throughout the cell cycle to give so-called replacement histones); (iii) tissue-specific, e.g. the H1t and H3t genes are expressed exlusively in the testis. There appears to be a total of 61 human histone genes which comprise several subfamilies. Most of the histone genes are found in two multifamily clusters on the short arm of chromosome 6, but genes on several other human chromosomes can specify identical copies of a particular histone subtype **(Fig.4)** .

Functionally similar genes

A large fraction of human genes are members of gene families where individual genes are closely related but not identical in sequence. In many such cases the genes are clustered and have arisen by tandem gene duplication, as in the case of the different members of each of the beta-globin and alpha globin gene clusters. Genes which encode clearly related products but which are located on different chromosomes are generally less related, as in the case of the alpha-globin and beta-globin genes.

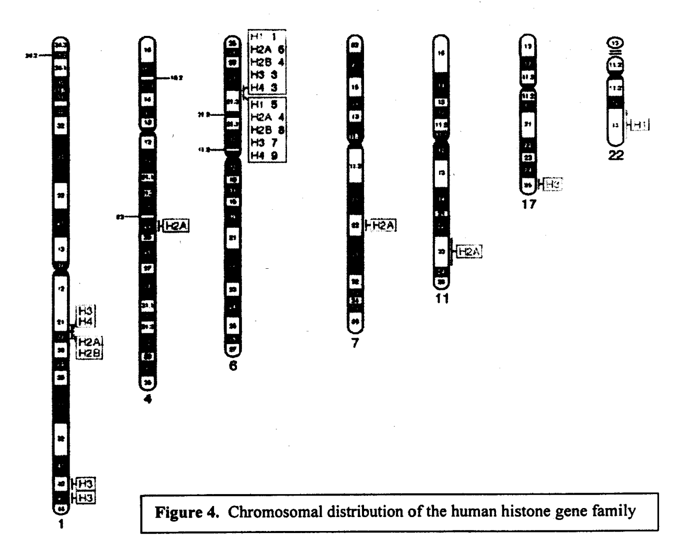


Figure 4. Chromosomal distribution of the human histone gene family

Functionally related genes

Some genes encode products which may not be so closely related in structure, but are clearly functionally related. The products may be subunits of the same protein or macromolecular structure, components of the same metabolic or developmental pathway, or may be required to specifically bind to each other as in the case of ligands and their relevant receptors. In almost all such cases, the genes are not clustered and are usually found on different chromosomes .

Size diversity

Genes in simple organisms such as bacteria are comparatively similar in size, and usually very short. By contrast, complex organisms such as mammals show wide variation in gene size, a feature found especially in human genes which can vary in length from hundreds of nucleotides to several megabases. The enormous size of some human genes means that transcription can be time-consuming. For example, the human dystrophin gene requires about 16 hours to be transcribed, and transcripts undergo splicing before transcription is completed.

As one would expect, there is a direct correlation between the size of a gene and the size of its product, but there are some striking anomalies. For example, apolipoprotein B has 4563 amino acids and is encoded by a 45 kb gene while the dystrophin gene is 2.4 Mb in length and encodes a product in muscle cells of 3685 amino acids.

Summary

The human genome is the term used to describe the total genetic information (DNA content) in human cells. It really comprises two genomes: nuclear (65,000–100,000 genes)and mitochondria (37 genes). Only 3% of the genome are coding sequences (genes).

Pseudogenes

They are truncated gene copies and gene fragments which are commonly found in multigene families. It is characterized by defective copies of the gene or it is a single exon (gene fragments).

Gene superfamilies

They are genes which encode products (proteins) that are functionally related. They show only very weak sequence homology over a large segment, and, there are some evidence for general common structural features.

Tandemly repeated noncoding DNA

Defined as blocks (or arrays) of tandemly repeated DNA sequences at a few or many different chromosomal locations. According to the size of the arrays of repeat units, (satellite, minisatellite and microsatellite.).

Satellite DNA composed of **very long** arrays of tandem repeats. DNA of this type is not transcribed and accounts for the bulk of the heterochromatic regions of the genome, specially in the region of the centromeres (pericentromeric heterochromatin).

Minisatellite: **Moderately long** arrays of tandemly repeated DNA sequences which are dispersed in the genome. They are not normally transcribed.

Hypervariable minisatellite DNA are **short tandem repeats.** Mainly found near the telomeres. The great majority them are not transcribed, except for elements occuring within noncoding intragenic sequences.

Genetic markers Hypervariable minisatellite have been characterized and used as **genetic markers**.

DNA fingerprinting: A single DNA probe contains the common sequence can hybridize to multiple minisatellite DNA on all chromosomes, resulting in individual-specific hybridization pattern.

Telomeric DNA
A major family of minisatellite DNA is found at the termini of chromosomes. These simple repeats are directly responsible for telomere function.

Microsatellite DNA: DNA families include **small** arrays of tandem repeats which are simple sequence interspersed throughout the genome. Mononucleotide repeats, dinucleotide (CA) repeats and three nucleotides repeats.

Interspersed repetitive non-coding DNA: The individual repeat dispersed at numerous locations in the genome, SINEs and LINEs.

SINEs (Short Interspersed Nuclear Elements): The most conspicuous human SINE is the Alu repeat family (so called because of early attempts at characterizing the sequence using the restriction nuclease Alu I).

LINEs (Long Interspersed Nuclear Elements): Human LINEs are exemplified by the LINE-1 or L1 element.

RNA-encoding gene families encode mature RNA molecules; **(1)** ribosomal RNA genes, **(2)** small nuclear RNA and **(3)** genes for transfer RNA.

Functionally similar genes
Are of three types:

Functionally identical genes Some genes on different chromosomes encode identical polypeptides e.g. histone gene subfamilies.

Functionally similar genes Gene families with individual closely related genes but they are not identical in their sequence. e.g. different members of the beta-globin and alpha globin gene.

Functionally related genes Some genes encode products which may not be so closely related in structure, but are clearly functionally related. The products may be subunits of the same protein or macromolecular structure.

Size diversity There is a direct correlation between the size of a gene and the size of its product, with some exceptions.

DNA (60%) **Single copy, Moderately repetitive** (30%) -**Highly repetitive** (10%) - There are numerous copies of the same sequence.

CHAPTER 2

STRUCTURE of DNA and RNA

Introduction:

Molecular genetics is primarily concerned with the inter-relationship between the information macromolecules DNA (deoxyribonucleic acid) (but in all cells genetic information is stored in DNA molecules) and RNA (ribonucleic acid) and how these molecules are used to synthesize polypeptides, the basic component of all proteins. The great majority of the RNA molecules are in turn used to specify the synthesis of polypeptides, either directly or by assisting at different stages in gene expression.

As the vast majority of gene expression is dedicated to polypeptide synthesis, proteins are the major functional end-points of the DNA and account for the majority cell structure. The term protein was derived from the Greek proteios, meaning 'of the first rank' and reflects the important roles of proteins in diverse cellular functions, as enzymes, receptors, storage proteins, transport proteins, transcription factors, signaling molecules, hormones, etc.

A- DNA (Deoxyribonuleic acid)

1. The structure of DNA is a double helix in which two DNA molecules (DNA strands) are held together by weak hydrogen bonds to form a DNA duplex **(Fig.5)**.

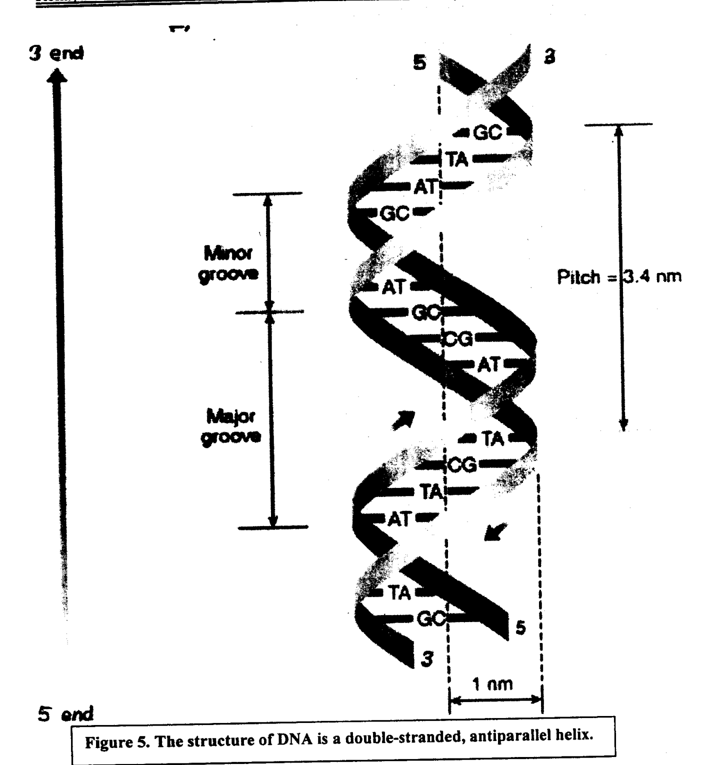


Figure 5. The structure of DNA is a double-stranded, antiparallel helix.

2. The linear backbone of a DNA molecule consists of alternating sugar residues and phosphate groups.

Figure 6. Hydrogen and phosphodiester bonds

3. The bond linking an individual sugar residue to the neighbouring sugar residues is a 3′, 5′-phosphodiester bond. This means that a phosphate group links carbon atom 3′ of a sugar to carbon atom 5′ of the neighbouring sugar. As the

phosphodiester bonds link carbon atoms number 3′ and number 5′ of successive sugar residues, one end of each DNA strand, the so-called 5′ end, will have a terminal sugar residue in which carbon atom number 5′ is not linked to a neighboring sugar residue. The other end is defined as the 3′ end because of a similar absence of phosphodiester bonding at carbon atom number 3′ of the terminal sugar residue **(Fig.6)**.

4. The two strands of a DNA duplex are said to be antiparallel because they always associate (anneal) in such a way that the 5′ → 3′ direction of one DNA strand is the opposite to that of its partner.

Nitrogenous bases

There are two types of nitrogenous bases

A- Purines are double ring compounds. In the DNA and RNA; purines found are adenine (A) and guanine (G) **(Fig.7 A)**.

B- Pyrimidines: are single ring structures **(Fig.7 B)**. They are of three types cytosine (C), thymine (T) and uracil (U). Cytosine and thymine are present in DNA and uracil and cytosine in the RNA.

1. A combination of deoxyrhibose sugar and nitrogenous base is called nucleoside and combination of a nucleoside and phosphate is known as nucleotide. There are 10 nucleotides per turn.
2. Union of several nucleotides together leads to formation of polypeptide chain. DNA molecule is a polymer which is composed of several thousand of nucleotide monomers.

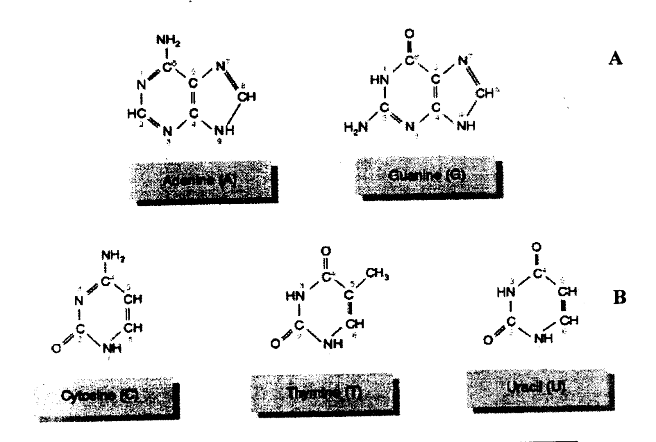


Figure 7. Structure of bases, nucleosides and nucleotides.

3. The total number of pyrimidines in one DNA molecule equal to the total number of purines . Thus A + G + C + T.

4. It is usual, therefore, to describe a DNA sequence by writing the sequence of bases of one strand only, and in the 5′ → 3′ direction. This is the direction of synthesis of new DNA molecules during DNA replication, and also of the nascent RNA strand produced during transcription.

Base pairing

According to **Watson-Crick rules**: The two strands of the DNA duplex adenine (A) specifically binds to thymine (T) by double hydrogen bonds and cytosine (C) specifically binds to guanine

(G) by triple hydrogen bonds. All hydrogen bonds are of week bonds which help in replication.**(Fig.8)**.

Figure 8. A-T base pairs have two connecting hydrogen bonds; G-C base pairs have three. hydrogen bonds

Classification of DNA

The DNA can be classified according to various characteria (1) number of base per turn (2) coiling pattern (3) location (4) structure (5) nucleotide sequence and (6) number of strands.

1-Number of base per turn

1- **B-DNA:** They have respectively 10 base pairs per turn. Under physiological conditions, most of the DNA in a bacterial or eukaryotic genome is of the B-DNA form in which each helical strand has a pitch (the distance occupied by a single turn of the helix) of 3.4 nm. The base pairs are perpendicular **(Fig.6)**.

2- **A-DNA:** This form found when the humidity of the sample is 75%. It has pitch 2.8nm and 11 base pairs per

turn of the helix. The base pairs are tilted, so the depth of deep groove is increased while the that of shallow grooves is reduced.

3- **C-DNA:** It is observed when the humidity of the sample is reduced to less than 66%. It has pitch of 3.1nm and 9.3 base pairs per turn. The base pair has negative tilt (-7.8).

4- **D-DNA:** This the rare form of the DNA. It has 8 base pair per turn, and the tilt is more negative than C-DNA.

5- **Z-DNA:** Which has 12 base pairs per turn and a pitch of 4.5nm. It so called because the sugar and phosphate bonds follow zigzag pattern. This DNA is associated with gene expression.

2-Coiling pattern

1- **Right handed**: Most of DNA molecules are right handed means that the coiling of the helix is in the right direction. A,B,C and D DNA forms all, are right handed.

2- **Left handed:** The Z DNA form is left handed.

3- location

1.**Nuclear DNA:** Which accounts for 99.9- 99.5% of the total DNA.

2.**Mitochondrial DNA** Which accounts for the remaining 0.0005%.

3.**Promiscuous DNA** this DNA sequences can move between organelles. It is recorded in some plants and in the yeast.

4- Structure

On the basis of structure there are two types of DNA;

1. **Linear DNA:** Which has thread like structure with both free ends and it is found in the eukaryotes.
2. **Circular DNA:** Which is ring or circular shape structure. It is found in bacteria and mitochondria

5- Repetition of base sequences

According the copies of nucleotide sequences the chromosomal DNA is classified into

1. **Single copy DNA:** In the genome some nucleotides or segments have only single copy per genome. Such DNA is called unique DNA. It is usually present higher in animals more than in plants. In human and mouse it represents about 70% of all genome but it accounts 22% of maize genome.
2. **Repetitive DNA:** The DNA which has small sequences of bases and these sequences are repeated several hundred times so it is known as repetitive DNA or satellite DNA.(discussed briefly latter).

6- Number of strands

According t the number of strands per DNA molecule there are two types of DNA:

1. **Double stranded DNA**: The DNA which has spirally arranged double strands and it was described by Watson and Crick.
2. **Single stranded DNA:** It is one helix and was recorded in very rare types of bacteriophage.

Complementarity:

Genetic information is encoded by the linear sequence of bases in the DNA strands (the primary structure). Consequently, two DNA strands of a DNA duplex are said to have complementary sequences and the sequence of bases of one DNA strand can readily be inferred if the DNA sequence of its complementary strand is already known.

However, when describing the sequence of a DNA region encompassing two neighboring bases (really a dinucleotide) on one DNA strand, it is usual to insert a ‘p' to denote a connecting phosphodiester bond e.g. CpG means that a cytosine is covalently linked to a neighboring guanosine on the same DNA strand, while a CG base pair means a cytosine on one DNA strand is hydrogen-bonded to a guanine on the complementary strand

Replication of DNA

Origins of replication: DNA replication is initiated at specific points, which have been termed origins of replication. The site of initiation differs from organism to organism. Starting from such an origin, the initiation of DNA replication results in a **replication fork**, where the parental DNA duplex bifurcates into two daughter DNA duplexes.

Unwound of DNA strand: The two DNA strands of each chromosome are unwound by a **helicase enzyme** and each DNA strand directs the synthesis of a complementary DNA strand to generate two daughter DNA duplexes, each of which is identical to the parent molecule.

During the process of DNA synthesis (DNA replication), the two strands of the parental DNA duplex are anti-parallel, but

act individually as templates for the synthesis of a complementary anti-parallel daughter strand. It follows that, the two daughter strands must run in opposite directions i.e. the direction of chain growth must be 5′ → 3′ for one daughter strand, the **leading strand,** but 3′ → 5′ for the other daughter strand, **the lagging strand. (Fig.9)**

Semi-conservative: As each daughter DNA duplex contains one strand from parent molecule and one newly synthesized DNA strand, the replication process is described as semi-conservative. **The leading strand** only will have a free 3′ hydroxyl group at the point of bifurcation. This will permit sequential addition of nucleotides and continuous elongation in the same direction in which the replication fork moves

The lagging strand However, synthesis of the lagging strand has to be accomplished as a progressive series of small fragments (typically 100–1000 nucleotides long), often referred to as **Okazaki fragments.**

Semi-discontinuous As only the leading strand is synthesized continuously, the synthesis of DNA strands is said to be semi-discontinuous. Each fragment of the lagging strand (Okazaki fragments) is synthesized in the 5′ → 3′ direction, which will be in the opposite direction to that in which the replication fork moves. Successively synthesized fragments are covalently joined at their ends using the enzyme DNA ligase so as to ensure chain growth in the direction of movement of the replication fork **(Fig.9).**

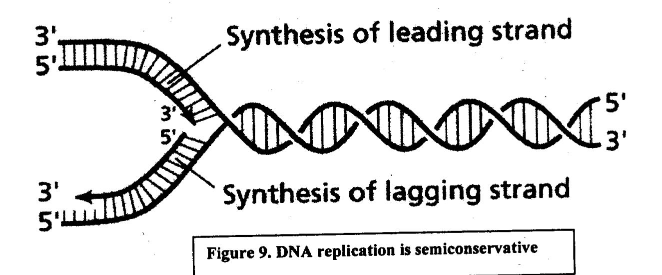


Figure 9. DNA replication is semiconservative

DNA polymerases The reactions catalyzed by DNA polymerases **(Fig.10)** involve addition, to the free 3′ hydroxyl group of the growing DNA chain, of a dNMP moiety provided by a dNTP precursor [the two distal phosphate residues of the dNTP - that is, the β and γ residues - are cleaved and the resulting pyrophosphate group (PP_i) discarded].

Five classes of mammalian DNA polymerases are known, including a polymerase that is dedicated to replication of the mitochondrial genome **(Table 1).**

In individual mammalian chromosomes, DNA replication proceeds bidirectionally, to form **replication bubbles** from multiple initiation points. The distance between adjacent replication origins is about 50–300 kb, a distance which may be significant in chromosome structure.

Table 1. The five classes of mammalian DNA polymerase

Class	α	β	γ	δ	ε
Location	Nuclear	Nuclear	Mitochondrial	Nuclear	Nuclear
Function	Synthesis and priming of lagging strand	DNA repair	Replicates mitochondrial DNA	Synthesis of leading strand	DNA repair
3′ → 5′ exonuclease?	No	No	Yes	Yes	Yes

At different origins, DNA replication is initiated at different times in the S phase of the cell cycle, but eventually neighboring replication bubbles will fuse. DNA replication is time-consuming: human cells in culture require about 8 hours to complete the process.

B- RNA (Ribonucleic acid)

On the basis of the functions, RNA is classified into two types, (1) genetic RNA, and (2) non-genetic RNA. The non genetic RNA is of three types, transfer RNA or soluble RNA, messenger RNA or template RNA and ribosomal RNA.

Genetic RNA

RNA which acts as DNA or the genetic material is referred to as genetic RNA. Such RNA is found in most of the plant viruses (TMV, HRV etc.), some animal viruses and certain bacteriophages. Genetic RNA may be single stranded or double stranded. This type of RNA is self replicative.

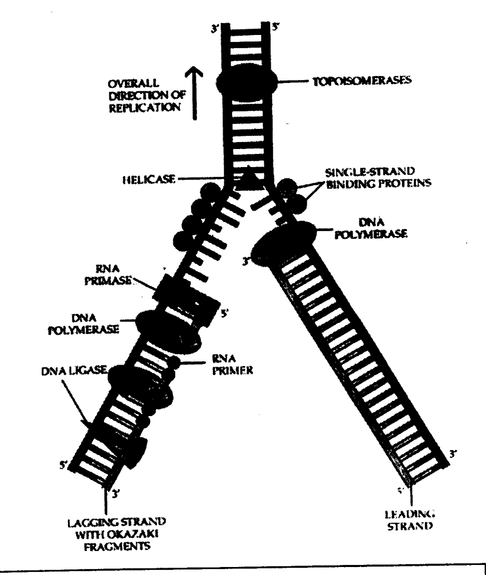


Figure 10. Asymmetry of strand synthesis during DNA replication

Non-Genetic RNA

RNA which does not act as genetic material is known as non genetic RNA. This is found in higher organisms where DNA is the genetic material. Such RNA is usually single stranded. This type of RNA does not have self replication property. Such RNA is synthesized from DNA template in the presence of DNA

dependent RNA polymerase enzyme. Thus, genetic RNA differs from non genetic RNA in several aspects.

♦ Messenger RNA (mRNA)

Since this RNA carries information for protein synthesis from DNA **(Fig.11)** to ribosomes (sites of protein synthesis), it is called messenger RNA. It constitutes about 5 – 10% of the total cellular RNA. The molecular weight of an average molecule of mRNA is 500.000. Its sedimentation coefficient is 8S. In bacteria, it is short lived. For example, in E. coli the average half life of some mRNA is about 2 minutes. However, in mammals it may live for many hours and even days. Usually each gene transcribes its own RNA. Hence, there are as many types of mRNA molecules as there are genes. There may be 1000 – 10.000 types of mRNA in a cell. These types of mRNA differ only in the sequence of their bases and in their length. When one gene (cistron) codes for one molecule of mRNA, it is known as monocistronic mRNA. On the other hand, when one mRNA molecule is coded by several cistrons, it is known as polycistronic or polygenic mRNA. New mRNA is synthesized during early cleavage on a DNA strand in the presence of RNA polymerase enzyme. Synthesis of mRNA differs from DNA replication in following three main aspects :

1. Ribose nucleotides are used instead of deoxyribose nucleotides.
2. Adenine pairs with uracil instead of thymine.
3. Only one strand of DNA leads to formation of mRNA molecule

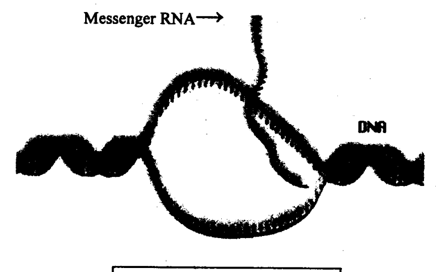


Figure 11. Messenger RNA

♦ **Ribosomal RNA (rRNA)**

The RNA which is found in ribosomes in the cytoplasm is called ribosomal RNA. It constitutes about 80% of the total cellular RNA. Its main features are given below :

1. Ribosomal RNA is more stable than mRNA.
1. Ribosomal RNA is synthesized from nucleolar DNA in eukaryotes and from a part of DNA in prokaryotes.
2. Synthesis of rRNA begins during gasturation and increases as embryo develops.
3. On the basis of molecular weight and sedimentation rate ribosomal RNA is of three types, (1) with molecular weight over a , million (21S-29S RNA). (2) with molecular weight below one million (12S-18S) and (3) with low molecular weight (5S RNA).

The function of ribosomal RNA is binding of mRNA and tRNA to ribosomes.

♦ Transfer RNA (tRNA)

Transfer RNA is also known as soluble RNA (sRNA). It constitutes about 10 – 15% of the total RNA of the cell. Its main features are given below :

1. It has a molecular weight of about 25.000 – 30.000 with sedimentation coefficient of 3.8S.
2. It contains 73 to 93 nucleotides. **(Fig.12).**
3. There are more than a hundred different types of tRNA per bacterial cell.
4. It is synthesized on a DNA template using small-section of the DNA molecule. It is synthesized at the end of cleavage.
5. Its main function is to carry various types of amino acids and attach them to mRNA template for synthesis of protein. There are 20 amino acids. Therefore, there should be at least 20 types of tRNA. However, tRNA is always more than 20 and in some cases. There are two types of tRNA which specify for one amino acid.

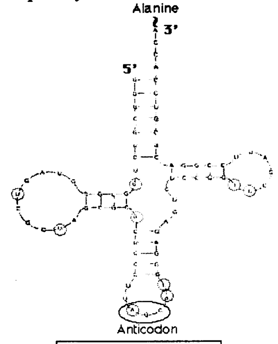


Figure 12. t RNA

Summary

Molecular genetics is primarily concerned with the inter-relationship between the information stored on the DNA, RNA and proteins. .

A- DNA

1. DNA is a antiparallel double helix formed of two DNA strands and held together by weak hydrogen bonds
2. The linear backbone of a DNA is sugar and phosphate..

Base pairs two types of nitrogenous bases

Pyrimidines: are single ring structures (C), thymine (T) and uracil (U). Purines are double ring compounds; adenine (A) and guanine (G) . (A) specifically binds (T) by double hydrogen bonds and (C) specifically binds (G) by triple hydrogen bonds.

1. nucleoside is deoxyribose sugar and nitrogenous base
2. nucleotide is nucleoside and phosphate
3. poly-peptide is the union of several nucleotides

Classification of DNA

The DNA can be classified according to various characteria (1) number of base per turn (2) coiling pattern (3) location (4) structure (5) nucleotide sequence and (6) number of strands.

Complementarity: The sequence of bases of one DNA strand can readily be inferred if the DNA sequence of its complementary strand is already known.

Replication of DNA

Origins of replication: DNA replication is initiated at specific points, results in a **replication fork**, where the DNA duplex bifurcates into two daughter DNA duplexes.

Unwound of DNA strand by a **helicase enzyme** and each DNA strand directs the synthesis of a complementary DNA strand to generate two daughter DNA duplexes,

Semi-conservative: Each daughter DNA duplex contains one strand from parent molecule and one newly synthesized DNA strand, the replication process is described as semi-conservative.

The leading strand only will have a free 3′ hydroxyl group at the point of bifurcation. So addition of nucleotides is continuous.

The lagging strand: Each fragment of the lagging strand (Okazaki fragments) is synthesized in the 5′ → 3′ direction, which will be in the opposite direction to replication fork movement. Fragments are joined using ligase enzyme

Semi-discontinuous: As only the leading strand is synthesized continuously, and lagging strand is discontinuous the synthesis of DNA strands is said to be semi-discontinuous.

B- RNA is one strand of Ribose nucleotides which use uracil instead of thymine). Types of RNA; (1) genetic RNA, and (2) non-genetic RNA. The non genetic RNA is of three types, messenger RNA ribosomal RNA and transfer RNA,.

- **Messenger RNA (mRNA)** Since this RNA carries information for protein synthesis from DNA to ribosomes (sites of protein synthesis), it is called messenger RNA. It constitutes 5 – 10% of the total cellular RNA, with molecular weight 500.000. and sedimentation coefficient is 8S. There may be 1000 – 10.000 types of mRNA in a cell. These types of mRNA differ only in the sequence of their bases and in their length.

- **Ribosomal RNA (rRNA** The RNA which is found in ribosomes in the cytoplasm is called ribosomal RNA. It constitutes about 80% of the total cellular RNA. It is more stable than mRNA.

- **Transfer RNA (tRNA)** Transfer RNA is also known as soluble RNA (sRNA). It constitutes about 10 – 15% of the total RNA of the cell. Its molecular weight of about 25.000 – 30.000 with sedimentation coefficient of 3.8S, contains 73 to 93 nucleotides There are more than a hundred different types of tRNA per bacterial. Its main function is to carry various types of amino acids and attach them to mRNA template for synthesis of protein. There are 20 amino acids. Therefore, there should be at least 20 types of tRNA.

CHAPTER 3



GENE AND GENE EXPRESSION

Introduction

Genes are the basic units of heredity in living cells. They consist of a length of DNA that contains instructions ("codes") for making a specific protein. Through these proteins, our genes influence almost everything about us, including how tall we will be, how we process foods, and how we respond to infections and medicines. Although most of our cells have the same genes, not all genes are active in every cell. Heart cells synthesize proteins required for that organ's structure and function; liver cells make liver proteins, and so on. In other words, not all the genes are "switched on" and expressed as proteins within every cell. Within an individual cell, the same genes may be switched on at some times and switched off at other times.

The flow of genetic information is almost exclusively one way: **DNA → RNA → protein**. The expression of genetic information in all cells is very largely a one way system: DNA specifies the synthesis of RNA and RNA specifies the synthesis of polypeptides, which subsequently form proteins. Because of its universality, the DNA → RNA → polypeptide (protein) flow of genetic information has been described as the **central dogma** of molecular biology.

Step 1: (transcription)

The synthesis of RNA using a DNA-dependent RNA polymerase, is described as transcription and occurs in the nucleus of eukaryotic cells and, to a limited extent, in mitochondria and chloroplasts.

Sense strand: Normally, only one of the two DNA strands acts as a template for RNA synthesis. During transcription, double-stranded DNA is unwound and the DNA strand which will act as a template for RNA synthesis forms a transient double-stranded RNA-DNA hybrid with the growing RNA chain. As the RNA transcript is complementary to this template strand, the transcript (RNA) has the same $5' \rightarrow 3'$ direction and base sequence (except that U replaces T) as the opposite, non-template strand of the double helix. For this reason the non-template strand is often called the sense strand **(Fig.13)**.

Anti-sense strand: The template strand is often called the anti-sense strand. In documenting gene sequences it is customary to show only the DNA sequence of the sense strand **(Fig.13)**. Orientation of sequences relative to a gene sequence is commonly dictated by the sense strand and by the direction of transcription (e.g. the 5′ end of a gene refers to sequences at the 5′ end of the sense strand, and sequences upstream or downstream of a gene refer to sequences which flank the gene at the 5′ or 3′ ends, respectively, of the sense strand).



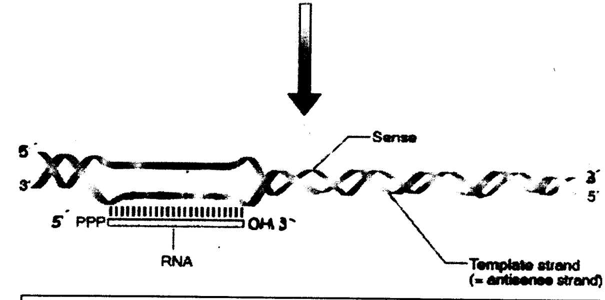


Figure 13. RNA is transcribed as a single strand which is complementary in base sequence to one strand (template strand) of a gene

Co-linearity principle: The expression of genetic information follows a co-linearity principle: the linear sequence of nucleotides in DNA is decoded to give a linear sequence of nucleotides in RNA which can be decoded in turn in groups of three nucleotides (codons) to give a linear sequence of amino acids in the polypeptide product.

Only a small fraction of the DNA in complex organisms is expressed to give a protein or RNA product

Transcription units: According to their needs, different cells transcribe different segments of the DNA (transcription units) which are discrete units, spaced

irregularly along the DNA sequence. However, the great majority of the cellular DNA is never transcribed in any cell **(Fig.14)**.Moreover, only a portion of the RNA made by transcription is translated into polypeptide. This is because:

1. Some transcription units are expressed to give an RNA molecules other than mRNA and so do not specify polypeptides directly, as in the case of ribosomal RNAs (rRNAs), tRNAs, and diverse small nuclear (sn) and cytoplasmic RNA molecules.
2. The primary transcript (initial transcription product) of those transcription units which do encode polypeptides is subject to RNA processing events. As a result, much of the initial RNA sequence is discarded to give a much smaller mRNA.
3. Only a central part of the mature mRNA is translated; sections of variable length at each end of the mRNA remain un-translated.
4. The fraction of coding DNA in the genomes of complex eukaryotes is rather small. This is partly a result of the non-coding nature of much of the sequence within genes entrons TATAA box,.. etc).
5. A considerable fraction of the genome of complex eukaryotes contains repeated sequences which are nonfunctional or which are not transcribed into RNA. These include defective copies of functional genes (pseudogenes and gene fragments), and highly repetitive non-coding DNA.

RNA synthesis is accomplished using an RNA polymerase enzyme, with DNA as a template and ATP, CTP, GTP and UTP as RNA precursors. The RNA is synthesized as a single strand, with the direction of transcription being 5′ → 3′. Chain elongation occurs by adding the appropriate ribonucleoside monophosphate residue (AMP, CMP, GMP or UMP) to the free 3′ hydroxyl group at the 3′ end of the growing RNA chain. Such nucleotides are derived by splitting a pyrophosphate residue (PPi) from the appropriate ribonucleoside triphosphate (rNTP) precursors. This means that the nucleotide at the extreme 5′ end (the initiator nucleotide) will differ from all others in the chain by carrying a 5′ triphosphate group.

In eukaryotic cells, three different RNA polymerase molecules are required to synthesize the different classes of RNA. The vast majority of cellular genes encode polypeptides and are transcribed by RNA polymerase II.

Regulation of transcription

Promotor: (cis-acting): They are short sequence elements in the immediate vicinity of a gene act as recognition signals for transcription factors to bind to the DNA in order to guide and activate the polymerase. They are often clustered upstream of the coding sequence of a gene, where they collectively constitute the promotor which can initiate transcription.

Trans-acting: They are transcription factors which are synthesized by genes remotely located, and migrate to their sites of action. In contrast, the promoter elements are **cis-acting**; their function is limited to the DNA duplex on which they reside. After a number of general transcription factors bind to the promoter region, an RNA polymerase binds to the transcription

factor complex and is activated to initiate the synthesis of RNA from a unique location.

In the case of genes which are actively transcribed by RNA polymerase II, either at a specific stage in the cell cycle (e.g. histones) or in specific cell types (e.g. β-globin), the promoter elements always include a TATA box, often TATAAA or a variant, at a position about 25 bp upstream (-25) from the transcriptional start site.

The promoters of many other genes, including housekeeping genes, lack TATA boxes but often have a GC box, containing variants of the consensus sequence GGGCGG.

Other common promoter elements include the CAAT box, often at about -80, which is usually the strongest determinant of promoter efficiency. Note, however, that the GC and CAAT boxes appear to be able to function in either orientation, although their sequences are asymmetrical.

In addition to general upstream transcription elements which are recognized by ubiquitous transcription factors, more specific recognition elements are known which are recognized by tissue-restricted transcription factors.

Enhancers comprise groups of cis-acting short sequence elements, which can enhance the transcriptional activity of specific eukaryotic genes. However, unlike promoter elements whose positions relative to the transcriptional initiation site are relatively constant, enhancers are located a variable, and often considerable, distance from the transcriptional start site, and their function is independent of their orientation.

They appear to bind gene regulatory proteins and, subsequently, the DNA between the promoter and enhancer loops out, allowing the proteins bound to the enhancer to interact with the transcription factors bound to the promoter, or with the RNA polymerase.

Silencers are equivalent regulatory elements which can inhibit the transcriptional activity of specific genes. Tissue-specific gene expression involves selective activation of specific genes, and regions of transcriptionally active chromatin adopt an open conformation

The DNA content of a specific type of eukaryotic cell, a myocyte for example, is virtually identical to that of a lymphocyte, hepatocyte or any other type of nucleated cell from the same organism. What makes the different cell types different is the pattern of genes which are expressed in the cell?. Some cells, particularly brain cells, express a large number of different genes. In many other cell types, a large fraction of the genes is transcriptionally inactive.

Housekeeping genes: Clearly, the genes that are expressed are the ones which define the functions of the cell. Some of these functions are common ones, which are essential for general cell functions and are specified by so-called housekeeping genes.

Tissue-specific gene expression: The expression of other genes may be largely restricted to a specific cell type (tissue-specific gene expression). Note, however, that even in the case of genes which show considerable tissue specificity in expression, some gene transcripts occur at very low levels in all cell types (illegitimate or ectopic transcription).

The distinction between transcriptionally active and inactive regions of DNA in a cell is reflected in the structure of the associated chromatin. Transcriptionally inactive chromatin generally adopts a highly condensed conformation and is often associated with regions of the genome which undergo late replication during S phase of the cell cycle. It is associated with tight binding by the histone H1 molecule. By contrast, transcriptionally active DNA adopts a more open conformation and is often replicated early in S phase. It is marked by relatively weak binding by histone H1 molecules and extensive acetylation of the four types of nucleosomal histones, i.e. histones H2A, H2B, H3 and H4. Additionally, in transcriptionally active chromatin the promoter regions of vertebrate genes are generally characterized by absence of methylated cytosines.

Step 2: RNA processing

The RNA transcript of most eukaryotic genes undergoes a series of processing reactions **(Fig.14)**.

1. Removal of unwanted internal segments
2. Rejoining of the remaining segments (RNA splicing).
3. In the case of RNA polymerase II transcripts, a specialized nucleotide linkage (7-methylguanosine triphosphate) is added to the 5′ end of the primary transcript (capping),
4. Adenylate (AMP) residues are sequentially added to the 3′ end of mRNA to form a poly(A) tail (polyadenylation).

The coding sequences of most vertebrate genes, both polypeptide-encoding genes and genes encoding RNA molecules other than mRNA, are split into segments (exons) which are separated by noncoding intervening sequences (interons). Transcription involves the production of RNA sequence

complementary to the entire length of the gene, encompassing both exons and introns. Often, the RNA transcript undergoes RNA splicing, a series of processing reactions whereby the intronic RNA segments are snipped out and discarded and the exonic RNA segments are joined end-to-end (spliced) to give a shorter RNA product.

Splicing mechanism:

1- **Splice junctions:** The identity of the nucleotide sequences at the exon/intron boundaries (splice junctions). Introns always start with GT (or really GU at the RNA level) and end with AG **(Fig.14)**.

Although the conserved GT (GU) and AG dinucleotides are crucially important for splicing.

2- **Sequences adjacent** to the GT and AG dinucleotides show a considerable degree of conservation.

3- **Branch site:** Which is usually located very close to the end of the intron, at most 40 nucleotides before the terminal AG dinucleotide.

Cleavage at the 5′ splice junction: Nucleolytic attack by the terminal G nucleotide of the splice donor site at the invariant A of the branch site to form a lariat-shaped structure; cleavage at the 3′ splice junction, leading to release of the intronic RNA as a lariat, and splicing of the exonic RNA segments. The above reactions are mediated by a large RNA-protein complex, the spliceosome, which consists of five types of snRNA (small nuclear RNA) and more than 50 proteins **(Fig.15)**.

In addition to RNA splicing, RNA polymerase II transcripts are subject to two additional RNA processing events.

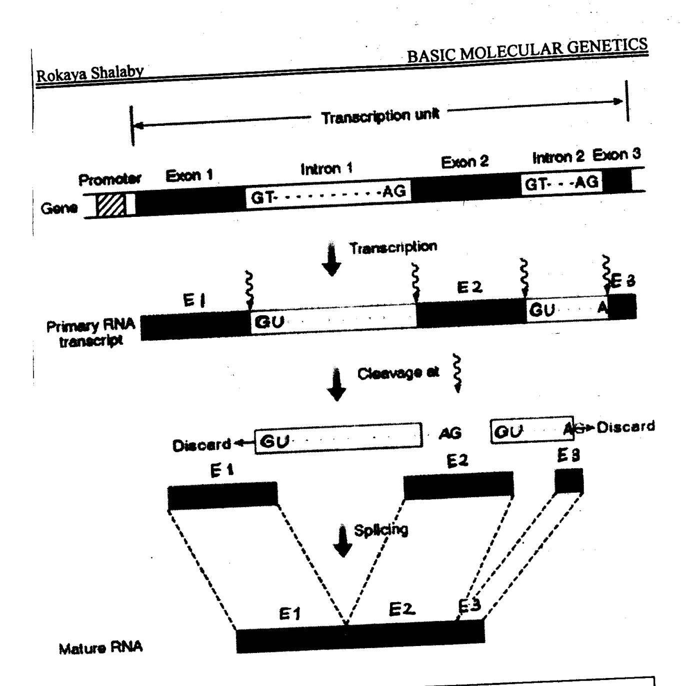


Figure 14. RNA splicing involves endonucleolytic cleavage and removal of intronic RNA segments and splicing of exonic RNA segments.

This occurs shortly after transcription. In the case of primary transcripts, which will be processed to give mRNA, a methylated nucleoside, 7-methylguanosine (m^7G) is linked to the first 5′ nucleotide of the RNA transcript by a special 5′-5′ phosphodiester bond. The 5′ end is said to be blocked or capped,

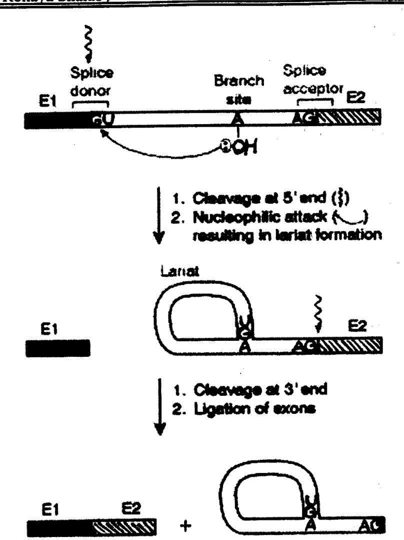


Figure 15. Mechanism of RNA splicing (GU-AG introns).
The nucleophilic attack involves the2-hydroxyl group attached to the conserved A at the branch site and the G of the conserved GU at the start of the intron, and results in a new covalent bond linking these two nucleotides to give a branched structure.

Capping

The cap has several possible functions:

1. to protect the transcript from 5′ → 3′ exonuclease attack (decapped mRNA molecules are rapidly degraded);
2. to facilitate transport from the nucleus to the cytoplasm.
3. to facilitate RNA splicing.
4. to play an important role in the attachment of the 40S subunit of the cytoplasmic ribosomes to the mRNA.

Polyadenylation

Transcription by both RNA polymerase I and III is known to stop after the enzyme recognizes a specific transcription termination site. However, identifying possible termination sites for transcription by RNA polymerase II is difficult because the 3′ ends of mRNA molecules are determined by a post-transcriptional cleavage reaction. The sequence AAUAAA is a major element that signals 3′ cleavage for the vast majority of polymerase II transcripts. Cleavage occurs at a specific site located 15–30 nucleotides downstream of the AAUAAA element **(Fig.16)**.

Following the cleavage point, transcription can continue for hundreds or thousands of nucleotides until termination occurs at one of several later sites. Once cleavage has occurred downstream of the AAUAAA element, about 200 adenylate (i.e. AMP) residues are sequentially added in mammalian cells by the enzyme poly(A) polymerase to form a poly(A) tail.

The poly(A) tail has several possible functions:

1. Facilitate transport of the mRNA to the cytoplasm.
2. Stabilize at least some of the mRNA molecules in the cytoplasm [shortening of poly (A) tracts is associated with mRNA degradation, but some mRNA species (e.g. actin mRNA) remain stable with little or no poly(A)].
3. It may facilitate translation by permitting enhanced recognition of the mRNA by the ribosomal machinery.

Step 3: Translation

The third step is translation. Polypeptide synthesis, is described as translation and occurs in the ribosomes, large RNA-protein

complexes which are found in the cytoplasm and also in mitochondria and chloroplasts. The RNA molecules which specify polypeptide are known as messenger RNA (mRNA). untranslated regions (5′ UTR; 3′ UTR), are originally copied from sequence derived from the 5′ and 3′ terminal exons and, like the 5′ cap and 3′ poly(A) tail, assist in binding and stabilizing the mRNA on the ribosomes where translation of the central segment occurs

Ribosomes are large RNA-protein complexes which provide a structural framework for polypeptide synthesis. They are composed of two subunits (with ultracentrifuge)

1. A large 60S subunit (contains three types of rRNA molecule: 28S rRNA, 5.8S rRNA and 5S rRNA, and about 50 ribosomal proteins).
2. A smaller 40S. (contains a single 18S rRNA and over 30 ribosomal proteins).
3. The RNA components are responsible for its catalytic function while the protein components are thought to enhance the function of the rRNA molecules, and a surprising number of them do not appear to be essential for ribosome function.

Translation in steps

1. **Ribosome binding**: The 40S ribosomal subunit initially recognizes the 5′ cap via the participation of proteins that specifically bind to the cap.
2. **Initiation**: The 40S scans along the mRNA until it encounters the initiation codon, which is almost always AUG (when it is embedded in a suitable sequence),

specifying methionine (a few cases are known where ACG, CUG or GUG are used instead) **(Fig.17)**.

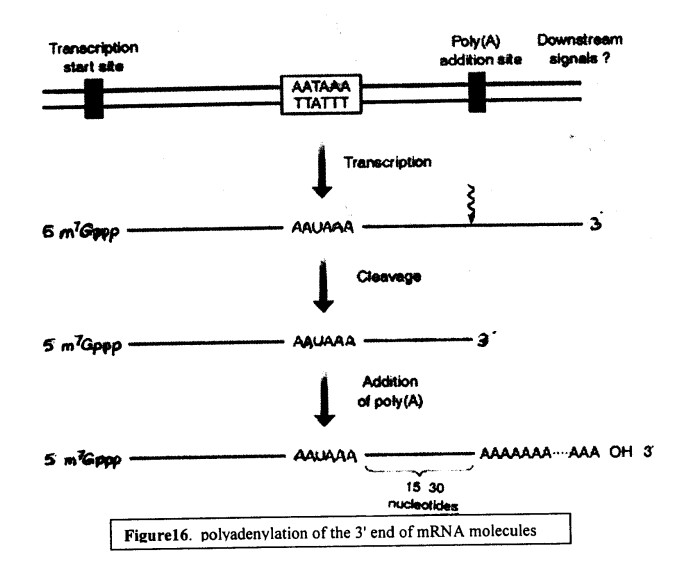


Figure16. polyadenylation of the 3' end of mRNA molecules

3. **Decoding**: The assembly of a new polypeptide from its constituent amino acids is governed by a triplet genetic code. Successive groups of three nucleotides in the linear mRNA sequence are decoded sequentially in order to specify individual amino acids. The decoding process is mediated by a collection of tRNA molecules, to each of which a specific amino acid has been covalently bound by

a specific amino acyl tRNA synthetase. Different tRNA molecules bind different amino acids.

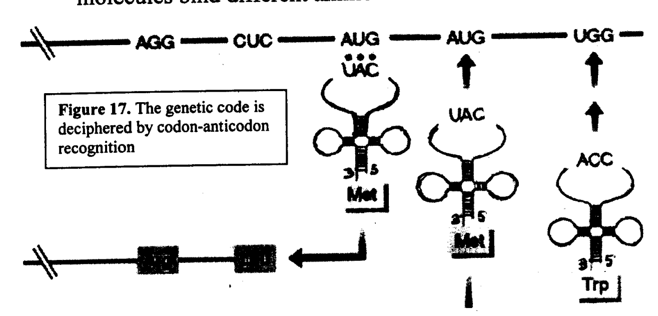


Figure 17. The genetic code is deciphered by codon-anticodon recognition

4. **Recognition**: Each tRNA has a specific trinucleotide sequence, called the anticodon, at a crucially important site located in the center of one arm of the tRNA. This site provides the necessary specificity to interpret the genetic code: the relevant codon of the mRNA molecule must be recognized via base-pairing with a suitably complementary anticodon of the appropriate tRNA molecule.
5. **Amino acids incorporation**: Subsequently, successive amino acids are incorporated into the growing polypeptide chain by a condensation reaction: the amino group of the incoming amino acid reacts with the carboxyl group of the last amino acid to be incorporated, resulting in a peptide bond between successive residues. This is catalyzed by a peptidyl transferase activity which resides in the RNA component of the large ribosomal subunit.
6. **Termination**: Translation continues until a termination codon is encountered (i.e. UAA, UAG or UGA). The

backbone of the primary translation product will therefore have at one end a methionine with a free amino group (the N-terminal end) and at the other end an amino acid with a free carboxyl group (the C-terminal end).

7. **Recruitment**: controlled by
    a. The predominant step in the control of translation is ribosome binding.
    b. In addition to the 5′ cap, the 5′ UTR (often < 200 bp)
    c. 3′ UTR (usually very much longer than the 5′ UTR)
    d. Several cis-acting elements.
    e. Few trans-acting factors which bind to these elements.
    f. It is possible that the 5′and 3′ UTR sequences interact to enhance translation. The 3′ UTR has a key role in translational regulation and signals for controlling translation, mRNA stability and localization have all been found in this region.

Step 4: Post-translational modifications

a- The addition of chemical groups which are attached covalently to the polypeptide chain at the translational and post-translational levels. This can involve simple chemical modification (hydroxylation, phosphorylation, etc.) of the side chains of single amino acids or the addition of different types of carbohydrate or lipid groups.

b- Addition of carbohydrate groups: Glycoproteins contain oligosaccharides which are covalently attached to the side chains of certain amino acids. Few proteins in the cytosol are glycosylated, that is have attached carbohydrate, and those that are carry a single sugar residue, N-acetylglucosamine, covalently

linked to a serine or threonine residue. By contrast, those proteins which are secreted from cells or exported to lysosomes, the Golgi apparatus or the plasma membrane are glycosylated. Oligosaccharide components of glycoproteins are largely preformed and added en bloc to polypeptides.

c- Protein modification by addition of lipid groups: Some proteins, notably membrane proteins, are modified by the addition of fatty acyl or prenyl groups which typically serve as membrane anchors.

d- Anchoring of a protein to the outer layer of the plasma membrane uses a different mechanism: The attachment of a glycosylphosphatidyl inositol (GPI) group.

Step 5: Post-translational cleavage:

1. The primary translation product may also undergo internal cleavage to generate a smaller mature product.
2. Occasionally, the initiating methionine is cleaved from the primary translation product, as during the synthesis of β-globin.
3. More substantial polypeptide cleavage is observed in the case of the maturation of many proteins, including plasma proteins, polypeptide hormones, neuropeptides, growth factors, etc.
4. All secreted polypeptides and also polypeptides which are transported across intracellular membranes are synthesized initially as precursors, in which a signal sequence (sometimes called a leader sequence) acts as a recognition signal for transport across cellular membranes. Thereafter

the signal peptide is cleaved from the main polypeptide and degraded.

5. Additionally, in some cases, a single mRNA molecule may specify more than one functional polypeptide chain as a result of proteolytic cleavage of a large precursor polypeptide.

Genetic code

Genetic code refers to the relationship between the sequence of bases in RNA and the sequence of amino acids in a polypeptide chain. In other words, the relationship between the four letter language of nucleotides and twenty letter language of amino acids is known as genetic code. DNA transcribes for mRNA and various base sequences of RNA code for 20 amino acids. There are four bases in RNA, A, G, U and C and there are 20 amino acids which take part in protein synthesis. Thus, more than 20 combinations are required to code for 20 amino acids and also for start and stop signals in the synthesis of polypeptide chain.

Singlet of base will have (4) = 4 combinations and doublet of base have (4) = 16 combinations which are insufficient to code for 20 amino acids. Therefore, triplet of RNA bases is required because it will give (4) = 64 combinations, which are sufficient to code for 20 amino acids as well as for start and stop signals.

Codon

The triplet sequence of RNA bases which codes for a particular amino acid is called a codon. There are 64 codons which constitute genetic code. In other words, genetic code is the set of all codons. The codons are of two types; sense codons and signal codons.

Sense Codons

Which code for amino acids are called sense codons. There are 61 sense condons which cods for 20 amino acids.

Signal Codons

Initiation codon: There is one codon, i.e., AUG which codes for start signal. This is known as start codon or initiation codon because it starts the synthesis of polypeptide chain. This codon also codes for amino acid methionine. In some cases valine (GUG) codes for start signal. In eukaryotes, the starting amino acid is methionine, while in prokaryotes it is N-formyl methionine.

Stop signal codon: There are 4 codons which code for signals. Signals are of two types, viz., start and stop signal. There are three codons (UAA, UAG and UGA) which code for stop signal. These codons are also known as stop codons or termination codons because they provide signal for the termination and release of polypeptide chain. Since stop signal codon do not code for any amino acid they were earlier called as nonsense codons. Signals of stop or termination codons are read by proteins called release factors. Stop signals are not read by tRNA molecules.

Anticodons

The base sequence of tRNA which pair with codon of mRNA during translation is called anticodon. Codons are written in 5'→3' diection, whereas anticodons are usually written in 3' →5' direction. In a codon, the first letter is at the 5' end while in anticodon the first letter is at the 3'end. There are 61 codons which code for amino acids. However, the number of tRNA molecules (anticodon) is always much lesser than codons in an

organism. Hence, anticodon of some tRNA molecules have to pair with more than one codon. There are four main differences between codons and anticodons.

Nature of genetic code

There are several important features or properties of a genetic code. The genetic code is (1) triplet, (2) universal, (3) commaless, (4) non-overlapping, (5) non-ambiguous, (6) degenerate or redundant and (7) has polarity. These features are briefly described below :

The Code is Triplet

The genetic code is triplet as already discussed. A singlet code has 4 codons and doublet code has 16 codons which are insufficient to code for 20 different amino acids which take part in the synthesis of protein. The triplet code has 64 (4) codons which are sufficient to code for 20 amino acids and also for start and stop signals in the synthesis of polypeptide chain. The triplet code was first suggested by Gamow in 1954. In a triplet code three RNA bases code for one amino acid **(Table 2)**.

The Code is Universal

The some genetic code is applicable to all forms of organisms from microbes to man. The experiments conducted by various workers with different types of organisms have proved beyond doubt that genetic code is universal. In other words, the same code is valid for all types of organisms. Exceptions to the universal applications of genetic code has been reported for mitochondrial genome.

The Code is Commaless

It is believed that the genetic cod is commaless. In other words, the codons are continuous and there are no demarcation lines between codons. Amino acids after the point of deletion as given below :

RNA basis:	UUU	(C)UC	GUA	UCC	ACC
Amino acids:	Phe	Leu	Val	Ser	Thr

The deletion of base C from leucine will change the genetic message in the following manner :

RNA basis:	UUU	UCG	UAU	CCA	CC
Amino acids:	Phe	Ser	Tyr	Pro	??

Experimental evidences also reveal that the code is commaless. Khorana and co-workers have also demonstrated that the genetic code is commaless.

The Cod is Non-Overlapping

These nucleotides or bases code for one amino acid. In a non-overlapping code, six bases will code for two amino acids. In a non-overlapping code, one base or letter is read only once. In overlapping code six nucleotides or basis will code for 4 amino acids, because each base is read three times.

The Cod is Non-Ambiguous

Out of 64 codons, 61 code for 20 different amino acids. However, non of the codons codes for more than 1 amino acid. In other words, each codon codes only for one amino acid. This clearly indicates that the genetic code is non ambiguous. In case

of ambiguous code, one codon should code for more than one amino acid. In the genetic code there is no ambiguity.

Table 2. The Genetic Code: The universal matching of codon to amino acids, used by all organisms.

	T	C	A	G
T	TTT Phenylalanine TTC Phenylalanine TTA Leucine TTG Leucine	TCT Serine TCC Serine TCA Serine TCG Serine	TAT Tyrosine TAC Tyrosine TAA Stop TAG Stop	TGT Cysteine TGC Cysteine TGA Stop TGG Tryptophan
C	CTT Leucine CTC Leucine CTA Leucine CTG Leucine	CCT Proline CCC Proline CCA Proline CCG Proline	CAT Histidine CAC Histidine CAA Glutamine CAG Glutamine	CGT Arginine CGC Arginine CGA Arginine CGG Arginine
A	ATT Isoleucine ATC Isoleucine ATA Isoleucine ATG Methionine	ACT Threonine ACC Threonine ACA Threonine ACG Threonine	AAT Asparagine AAC Asparagine AAA Lysine AAG Lysine	AGT Serineine AGC Serineine AGA Arginine AGG Arginine
G	GTT Valine GTC Valine GTA Valine GTG Valine	GCT Alanine GCC Alanine GCA Alanine GCG Alanine	GAT Aspartate GAC Aspartate GAA Glutamate GAG Glutamate	GGT Glycine GGC Glycine GGA Glycine GGG Glycine

The Code is Degenerate

In most of the cases several codons code for the same amino acid. Only two amino acid, (tryptophan and methionine) are coded by one codon each. Nine amino acid are coded by two codons each, one amino acid (Isoleucine) by three codons, five amino acids by 4 codons each, and 3 amino acids by six codons each. This multiple system of coding is known as degenerate or the redundant code system. Such system provides a protection to the organism against many harmful mutations, because if one base of a codon is mutated, there are other codons which will

code for the same amino acid and there will be no alteration in the polypeptide chain.

The redundancy or degeneracy of code is not random except for serine, leucine and arginine. All codons coding for same amino acid are in the same box (except above three). Thus, the first two letters are GC in all the four codons of alanine are GC and GU in the four codons of valine.

Wobble hypothesis:. The wobble hypothesis states that pairing of codon and anticodon follows the normal A-U and G-C rules for the first two base positions in a codon, but that exceptional 'wobbles' occur at the third position and G-U base pairs are also admitted **(Table 3)**.

Table 3. Codon-anticodon pairing admits relaxed base-pairing (wobbles) at the third base position of codons

Base at 5′ end of tRNA anticodon	Base recognized at 3′ end of mRNA codon
A	U only
C	G only
G	C or U
U	A or G

Summary

The flow of genetic information is almost exclusively one way: **DNA → RNA → protein. This need three steps**

Step 1: (transcription),

The synthesis of RNA using a DNA-dependent RNA polymerase, **Sense strand,** strand that acts as a template for RNA synthesis. RNA transcript has the same 5′ → 3′ direction as the opposite, strand. For this reason the non-template strand is called the sense strand. **Anti-sense strand.** is The template strand.

Co-linearity principle: the linear sequence of nucleotides in DNA is decoded to give a linear sequence of nucleotides in RNA which can be decoded in turn in groups of three nucleotides (codons) to give a linear sequence of amino acids in the polypeptide product.

Transcription units different cells transcribe different segments of the DNA which are discrete units (genes)

Regulation of transcription by (1)Promotor: **(cis-acting)** they are short sequence elements immediately next to the gene act as recognition signals for transcription factors to activate the polymerase. **(2) Trans-acting:** They are transcription factors which are synthesized by genes remotely located, and migrate to their sites of action. **(3)Enhancers** comprise groups of cis-acting short sequence elements, which can enhance the transcriptional activity of specific genes. and **(4) Silencers are** equivalent regulatory elements which can inhibit the transcriptional activity of specific genes.

Step 2: RNA processing

(1) Removal of unwanted internal segments (2) RNA splicing (3)capping (4)polyadenylation.

The poly(A) tail has several possible functions:

1. Facilitates transport of the mRNA to the cytoplasm.
2. Stabilizes at least some of the mRNA molecules in the cytoplasm [shortening of poly (A) tracts is associated with mRNA degradation, but some mRNA species (e.g. actin mRNA) remain stable with little or no poly(A)];
3. It may facilitate translation by permitting enhanced recognition of the mRNA by the ribosomal machinery.

Step 3: Translation :

Translation is the process whereby mRNA is decoded on ribosomes to specify the synthesis of polypeptides. Only the central segment of a typical eukaryotic mRNA molecule is translated to specify the synthesis of a polypeptide.

Translation in steps

(1) Ribosome binding: (2) Initiation: when ribosome fined the initiation codon, (AUG). (3) Decoding: Successive groups of three nucleotides in the linear mRNA sequence are decoded sequentially in order to specify individual amino acids. (4) Incorporation of amino acids into the growing polypeptide and (5)Termination: by one of termination codon (i.e. UAA, UAG or UGA).

The genetic code: The genetic code is a three-letter code. Certain amino acids are specified by a single codon; others, , are specified by six codons.

Anticodons: The base sequence of tRNA which pair with codon of mRNA during translation is called anticodon.

Nature of genetic code

There are several important features or properties of a genetic code. The genetic code is (1) triplet, (2) universal, (3) commaless, (4) non-overlapping, (5) non-ambiguous, (6) degenerate or redundant and (7) has polarity.

CHAPTER 4



MUTATIONS

Introduction

As in other genomes, the DNA of the human genome is not a static entity. Instead, it is subject to a variety of different types of heritable change (mutation). New mutations arise in single individuals, in somatic cells or in the germline. If a germline mutation does not seriously impair an individual's ability to have offspring he can transmit the mutation to other members of the (sexual) population. Mutations are the raw fuel that drives evolution (polymorphism), but they can also be pathogenic.

Polymorphism

In genetics, the coexistence of several distinctly different types in a population (groups of animals of one species). Examples: different blood groups in humans, different colour forms in some butterflies, and snail shell size, length, shape and colour. Allelic sequence variation is traditionally described as a DNA polymorphism, if more than one variant (allele) at a locus occurs in a population with a frequency greater than 0.01. The mean heterozygosity for human genomic DNA is thought to be of the order of 0.001- 0.004 (i.e. approximately 1:250 to 1:1000 bases are different between allelic sequences). Certain genes, notably some HLA genes, are exceptionally polymorphic and alleles can show very substantial sequence divergence.

Because mutation or polymorphism rates are comparatively low the vast majority of the differences between

allelic sequences within an individual are inherited, rather than resulting from de novo mutations.

Effects of mutations.

They can be the direct cause of a phenotypic abnormality or they can result in increased susceptibility to disease. The usually low level of mutation may therefore be viewed as a balance between permitting occasional evolution at the expense of causing disease or death in a proportion of the members of a species. In many cases, a deleterious gene mutation in a somatic cell may cause death for that single cell, but will not have consequences for other cells. However, in some cases, the mutation may lead to an inappropriate continuation of cell division, causing cancer.

Origin of mutations

Normally, most mutations arise as copying errors during DNA replication because DNA polymerases, like all enzymes, are error-prone. The error rate of a DNA polymerase is significantly reduced by having a subunit of the polymerase which has a proofreading function. However, the size of the human genome makes huge demands on the fidelity of any DNA polymerase: a sequence of 3 billion nucleotides needs to be replicated accurately every single time a human cell divides.

Sources of mutations

Spontaneous: DNA is subject to significant spontaneous chemical attack in the cell. For example, every day approximately 5000 adenines or guanines are lost from the DNA of each nucleated human cell by depurination (the N-glycosidic bond linking the purine residue to the carbon 1 of the deoxyribose is hydrolyzed and the purine is replaced by a

hydroxyl group at carbon 1). However, under normal circumstances the greatest source of mutations is from endogenous mutation, notably spontaneous errors in DNA replication and repair.

During an average human lifetime there are an estimated 10^{17} cell divisions: about 2×10^{14} divisions are required to generate the approximately 10^{14} cells in the adult, and additional mitoses are required to permit cell renewal in the case of certain cell types, notably epithelial cells. As each cell division requires the incorporation of 6×10^{9} new nucleotides, error-free DNA replication in an average lifetime would require a DNA replication-repair process with an accuracy great enough so that the correct nucleotide was inserted on the growing DNA strands on each of about 6×10^{26} occasions. Such a level of DNA replication fidelity is impossible to sustain; indeed, the observed fidelity of replication of DNA polymerases is very much less than this and uncorrected replication errors occur with a frequency of about 10^{-9}-10^{-11} per incorporated nucleotide. Thus, during the approximately 10^{16} mitoses undergone in an average human lifetime, each gene will be a locus for about 10^{8}-10^{10} mutations (but for any one gene, only a tiny minority of cells will carry a mutation).

Induced: Mutations can be induced in our DNA by exposure to a variety of mutagens occurring in our external environment or to mutagens generated in the intracellular environment. In the case of radiation-induced mutation, for example, the normal germline mutation rate for hypervariable minisatellite loci was doubled as a consequence of heavy exposure to the radiaoctive fallout from the **Chernobyl accident**.

DNA is also can be damaged by exposure to natural ionizing radiation and to reactive metabolites.

Large-scale mutations:

They are chromosome abnormalities which involve loss or gain of chromosomes or breakage and rejoining of chromatids and lead to phenotype abnormality e.g. Dawn syndrome.

Small-scale mutations

They can be grouped into different mutation classes and can also be categorized on the basis of whether they involve a single DNA sequence (simple mutations) or whether they involve exchanges between two allelic or nonallelic sequences (sequence rpeat mutation).

A- Simple mutations

Three classes of small-scale simple mutations can be distinguished:

- **Base substitutions** involve replacement of usually a single base. Base substitutions are among the most common mutations and can be grouped into two classes:

  Transitions are substitutions of a pyrimidine (C ↔ T) by a pyrimidine, or of a purine (A ↔ G) by a purine.

  Transversions are substitutions of a pyrimidine by a purine or of a purine by a pyrimidine

When one base is substituted by another, there are always two possible choices for transversion, but only one choice for a transition. For example, the base adenine can undergo two possible transversions (to cytosine or to thymine) but only one

transition (to guanine;). So it is expected that transversions is twice as frequent as transitions.

The different classes of base substitution show differential tendencies to be located at the first, second or third base positions of codons. Because of the design of the genetic code, different degrees of degeneracy characterize different sites. According to base positions in codons simple mutations can be grouped into three classes:

Non-degenerate sites, they include the first base position of all but eight codons, the second base position of all codons and the third base position of two codons, AUG and UGG. The base substitution rate at non-degenerate sites is very low, consistent with a strong selection pressure to avoid amino acid changes **(Fig.18)**.

Fourfold degenerate sites are found at the third base position of several codons. The substitution rate at fourfold sites is very similar to that within introns and pseudogenes,

Twofold degenerate sites are often found at the third base positions of codons, but also at the first base position in eight codons. As expected, the substitution rate for twofold degenerate sites is intermediate: only one out of the three possible substitutions, a transition, maintains the same amino acid. For example, at the third base position of the glutamate codon GAA, a transition A G is silent, while the two transversions (A C; A T) result in replacement by a closely similar amino acid, aspartate

The design of the genetic code and the degree to which one amino acid is functionally similar to another affect the

relative mutabilities of individual amino acids. Certain amino acids may play key roles which cannot be substituted easily by

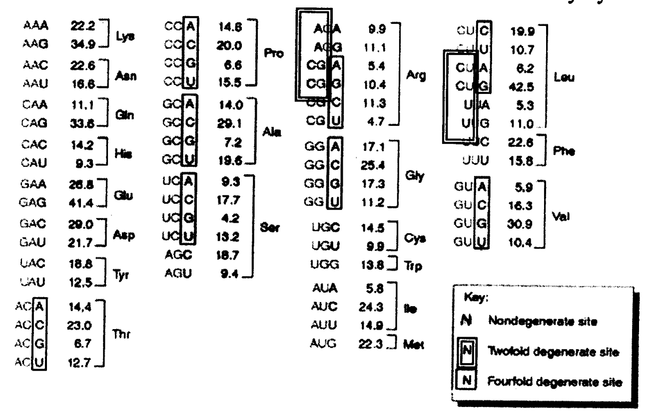


Figure 18. Codon frequencies in human genes and locations of nondegenerate, two- and fourfold degenerate sites. Note that eight of the 61 first base positions are twofold degenerate.

others. For example, cysteine is often involved in disulfide bonding which can play an important role in establishing the conformation of a polypeptide. As no other amino acid has a side chain with a sulfhydryl group, there is strong selection pressure to conserve cysteine residues at many locations, and cysteine is among the least mutable of the amino acids. In contrast, certain other amino acids such as serine and threonine have very similar side chains, and substitutions at both the first base position of codons (<u>A</u>CX <u>U</u>CX; where X = any nucleotide) and second base

positions (A<u>C</u>Py A<u>G</u>Py; where Py = pyrimidine) can result in serine threonine substitutions. Presumably as a result, serine and threonine are among the most mutable of the amino acids

- **Deletions** - one or more nucleotides are eliminated from a sequence.
- **Insertions** - one or more nucleotides are inserted into a sequence. In rare cases this involves transposition from another locus.
    1. **Copy transposition** involves a sequence from one locus being replicated and the copy inserted into another locus.
    2. **Non-copy transposition** involves simple transposition of a DNA sequence from one locus to another. In human and mammalian genomes, noncopy transposition is very rare.

Mutations in coding and noncoding DNA

Many mutations are generated essentially randomly in the DNA of individuals. As a result, coding DNA and noncoding DNA are about equally susceptible to mutation. Clearly, however, the major consequences of mutation are largely restricted to the approximately 3% of the DNA in the human genome which is coding DNA. Mutations which occur in this component of the genome are of two types:

1. **Silent (synonymous) mutations** do not change the sequence of the gene product. Silent mutations are thought to be effectively neutral mutations (conferring no advantage or disadvantage to the organism in whose

genome they arise). Silent mutation can result in certain diseases **(Fig.19)**.

2. **Non-synonymous mutations** result in an altered sequence in a polypeptide or functional RNA:. One or more components of the sequence are altered or eliminated, or an additional sequence is inserted into the product. Non-synonymous mutations can be grouped into three classes, depending on their effect:

a. **Deleterious effect**, those having a deleterious effect on gene expression and so can result in disease or death. However, the frequency of such mutation in the population is very much reduced because of natural selection.

b. **Beneficial effect**, those with a beneficial effect (e.g. improved gene function or gene-gene interaction).

c. **No effect**, those with no effect.

Hot spots of pathogenic mutations

Pathogenic mutations can occur at three types of DNA sequence at a gene locus.

1- **The coding sequence of the gene**, this is where the great majority of recorded pathogenic mutations have been identified. Those due to nucleotide substitution are, in the vast majority of cases, nonsynonymous substitutions and mostly occur at first and second base positions of codons. Because of its relatively high mutability, the CpG dinucleotide is often located at hotspots for pathogenic mutation in coding DNA. Other hotspots include tandem repeats within coding DNA.

2- Intragenic noncoding sequences, this is restricted to sequences which are necessary for correct expression of the gene, such as important intronic elements, notably the highly conserved GT and AG dinucleotides at the ends of introns.

3- Regulatory sequences outside exons, most mutations located in regulatory sequences have been identified in conserved elements located just upstream of the first exon, notably promoter elements. In addition, other more distantly located regulatory elements may be sites of pathological mutation.

Factors govern mutations expression

The degree to which a pathogenic mutation results in an aberrant phenotype depends on several factors:

1. The location of the mutation within the gene. For example, if gene dosage is carefully regulated, over-expression of that gene may cause an abnormal phenotype. Also ectopic expression, (expression in tissues where the gene is not normally expressed) may also be harmful.
2. The presence of a single normal allele may be sufficient to maintain a clinically normal phenotype (as in recessively inherited disorders), or a milder phenotype when compared with that of mutant homozygotes, as in dominantly inherited disorders where the mutation is a simple loss of function mutation.
3. The degree to which expression of a mutant phenotype is influenced by other gene products. The same mutant allele can have different phenotypic effects on different genetic backgrounds, depending on particular alleles at other gene loci (modifier genes).

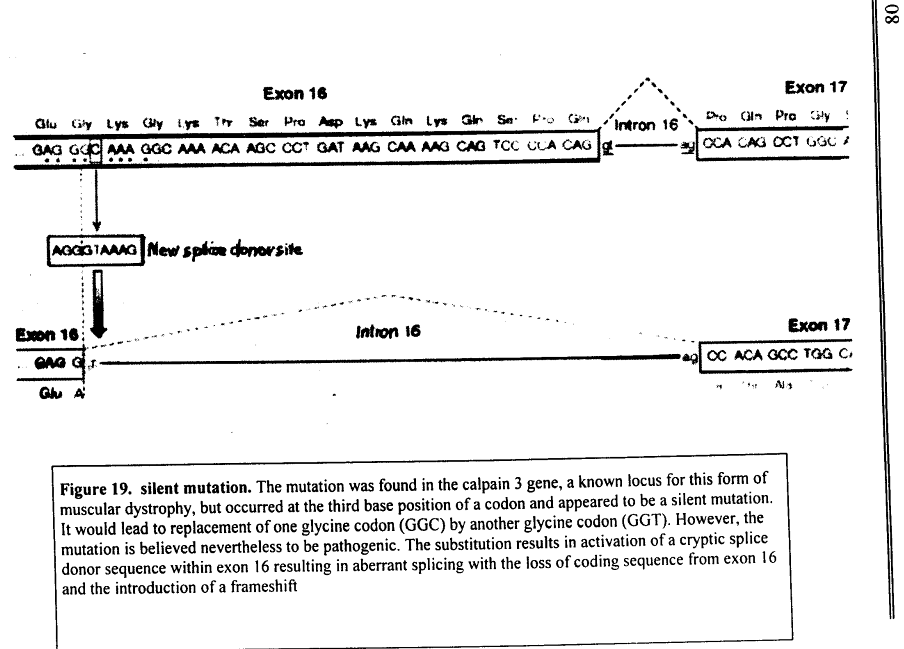


Figure 19. silent mutation. The mutation was found in the calpain 3 gene, a known locus for this form of muscular dystrophy, but occurred at the third base position of a codon and appeared to be a silent mutation. It would lead to replacement of one glycine codon (GGC) by another glycine codon (GGT). However, the mutation is believed nevertheless to be pathogenic. The substitution results in activation of a cryptic splice donor sequence within exon 16 resulting in aberrant splicing with the loss of coding sequence from exon 16 and the introduction of a frameshift

4. The proportion and nature of cells in which the mutant gene is present. Generally, mutations which are present in all cells of an individual (inherited mutations) or in many cells (somatic mutations acquired in very early development) are likely to have more profound effect than those present in a few cells (somatic mutations which arise at much later stages) or in cell types where the relevant gene is not expressed. Cancers, however, arise from unregulated division of cells produced from a single original mutant cell.
5. The parental origin of the mutation. This is only known to be important in the case of the few genes which are imprinted.

Gene conversion

This describes a nonreciprocal transfer of sequence information between a pair of nonallelic DNA sequences (interlocus gene conversion) or allelic sequences (interallelic gene conversion). The donor, remains unchanged. while the acceptor, is changed because some or all its sequence is replaced by a sequence copied from the donor sequence. The sequence exchange is therefore in one direction; the acceptor sequence is modified by the donor sequence, but not the other way round **(Fig.20)**.

B- Sequence repeats mutations

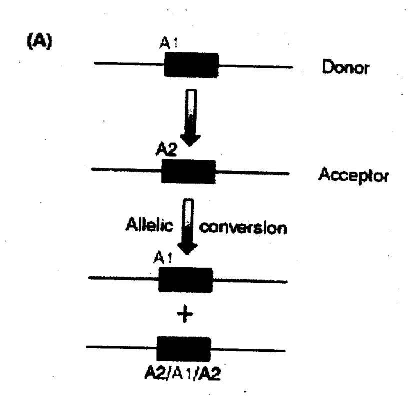
In addition to very frequent simple mutations, there are several mutation classes which involve sequence exchange between allelic or non-allelic sequences, often involving repeated sequences. For example, tandemly repetitive DNA is prone to deletion/insertion polymorphism whereby different alleles vary in the number of integral copies of the tandem repeat. Such

variable number of tandem repeat (VNTR) polymorphisms can occur in the case of repeated units that are very short (microsatellites); intermediate (minisatellites) or large repeats.

Genetic mechanisms for VNTR polymorphism.

Slipped strand mispairing at short tandem repeats (microsatellites). There is considerable variation in the germline mutation rates at microsatellite loci, ranging from an undetectable level up to about 8×10^{-3}. Novel length alleles at (CA)/ (TG) microsatellites and at tetranucleotide marker loci are known to be formed without exchange of flanking markers. This means that they are not generated by unequal crossover**(Fig.21)**. In addition to mispairing between tandem repeats, slippage replication has been envisaged to generate large deletions and duplications by mispairing between noncontiguous repeats and has been suggested to be a major mechanism genome evolution. The pathogenic potential of short tandem repeats is considerable

Unequal crossover Homologous recombination describes recombination (crossover) occurring at meiosis. It usually involves breakage of nonsister chromatids of a pair of homologs and rejoining of the fragments to generate new recombinant strands. Often the sequences at which crossover takes place show very considerable sequence homology which presumably stabilizes mispairing of the chromosomes. Because crossover occurs between mispaired nonsister chromatids, the exchange results in a deletion on one of the participating chromatids and an insertion on the other. Unequal crossover is also expected to occur comparatively frequently in complex satellite DNA repeats and at tandemly repeated gene loci. In the latter case, unequal crossover is known to generate pathogenic deletions at some loci.



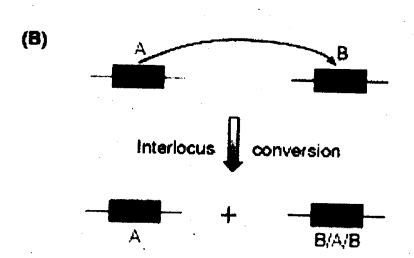


Figure 20. Gene conversion : (A)Interallelic gene conversion (B)Interlocus gene conversion

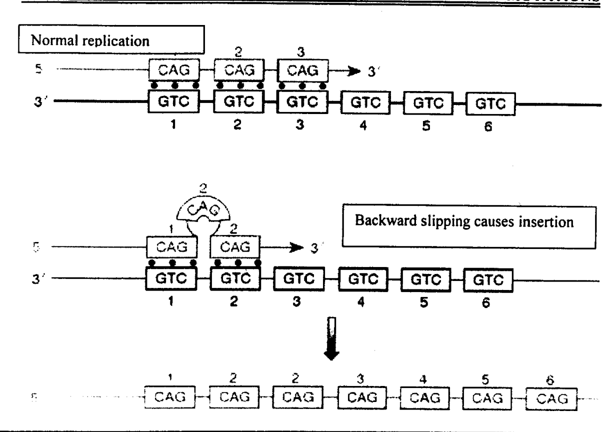


Figure 21 . Slipped strand mispairing during DNA replication causes insertions.

Unequal sister chromatid exchange. Sister chromatid exchange is an analogous type of sequence exchange involving breakage of individual sister chromatids and rejoining fragments in mitosis that initially were on different chromatids of the same chromosome **(Fig.22)** is thought to be a major mechanism underlying VNTR polymorphism. If chromosome breakage and rejoining occurs while the chromatids are mispaired, an insertion or deletion of an integral number of repeat units will result. Such exchanges can also lead to concerted evolution by causing a particular variant to spread through an array of tandem repeats, resulting in homogenization of the repeat units.

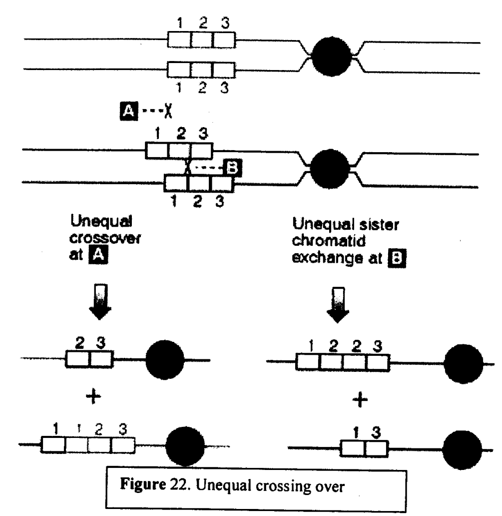


Figure 22. Unequal crossing over

Mitochondrial hotspots

Because of the very large size of the human nuclear genome, most mutations occur in nuclear DNA sequences. By comparison, the mitochondrial genome is a small target for mutation (about 1/200,000 of the size of the nuclear genome). Unlike nuclear genes, mitochondrial genes are present in numerous copies (there are thousands of copies of the mtDNA molecule in each human somatic cell; some cells, such as brain and muscle cells, have particularly high oxidative phosphorylation requirements and so more mitochondria). The mtDNA is inherited from the maternal oocyte. The mutation in mitochondrial DNA must arise in a single mtDNA molecule, so

the mutation rate correspondingly suspected to be low. On these grounds, you could anticipate that the proportion of clinical disease due to pathogenic mutation in the mitochondrial genome should be extremely low. Controversially , the frequency of 'mitochondrial disorders' is rather high and the mitochondrial genome can be considered to be a mutation hotspot.

Factors explain this apparent paradox:

1. The mitochondrial genome has a much higher percentage of coding DNA (93%) than found in the nuclear genome (3%). When this is taken into consideration, there is still a large imbalance: about 100 Mb of coding DNA in the nuclear genome but only 15.4 kb of coding sequence in the mitochondrial genome, giving a target ratio of 6000:1 in favor of the nuclear genome.

2. The mitochondrial genome is much more prone to nucleotide change than the nuclear genome. (mtDNA instability). The high instability of mtDNA has been postulated to result from several factors: (1)The high rate of production of reactive oxygen species by the respiratory reactions causes oxidative damage to mtDNA, (2)Unlike nuclear DNA, mtDNA is not protected by histones. (3) The mtDNA also has to undergo many more rounds of replication than chromosomal DNA. (4) Although several well-characterized mtDNA repair systems are now known, some frequent mutations cannot be repaired, including thymidine dimers.

3. Mutations have been reported to be fixed in the mitochondrial genomes of animal cells at a rate which is

about 10 times greater than occurring in equivalent sequences in the nuclear genome.

Mutations of mRNA

Often such mutations occur at the invariant GT and AG dinucleotides located respectively at the start of an intron (splice donor) or at its end (splice acceptor). **(Fig.23)** Flanking these important signals, however, are other conserved sequence elements which, if mutated, can also cause aberrant splicing. Mutations which alter such sequences can have different consequences:

1. Failure of splicing causing intron retention. The introduction of intronic sequence into the coding sequence of a mature mRNA will introduce additional amino acids and may cause a frameshift mutation.
2. Exon skipping. Often, the exclusion of an exon has a profound effect on gene expression: it may result in a frameshift, an unstable RNA transcript, or a nonfunctional polypeptide because of a loss of a critical group of amino acids.
3. Several different classes of mutation can introduce a premature termination codon (chain terminating mutations). Nonsense mutations produce a premature termination codon simply by substituting a normal codonwith a stop codon. Frameshifting insertions and deletions usually also introduce a premature termination codon not too far downstream of the mutation site.
4. Unstable mRNA. This is the most frequent consequence. A mRNA carrying a premature codon is usually rapidly degraded in vivo by a form of RNA surveillance known as

nonsense-mediated mRNA decay. This can avoid the potentially lethal consequences of producing a truncated polypeptide which could interfere with vital cell functions

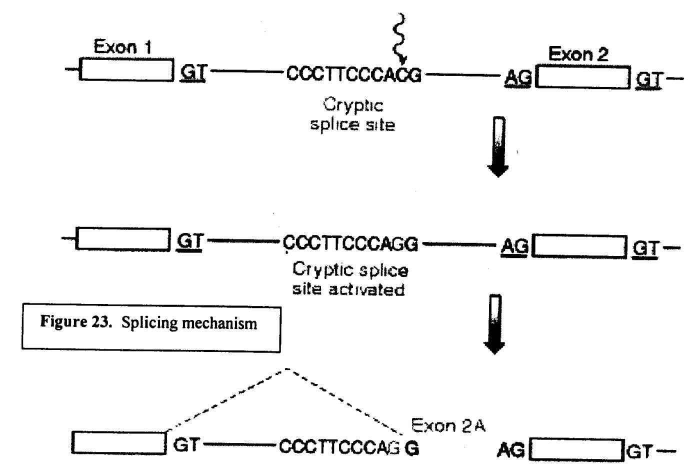


Figure 23. Splicing mechanism

.Frameshift mutations

The mutation which arise due to additions or additions of nucleotides in mRNA are known as frameshift mutations, because the normal reading frame of base triplets (codons) is altered as a consequence of such mutations .The addition or deletion of nucleotides occur in numbers other than three or multiple of three. The reading frame in such case is shifted from the point of addition or deletion onwards. The addition or deletion of base pairs takes place in interstitial or intercalary

position. Sometimes, addition and deletion take place at the same position, they are known as double frameshifts. Such changes may restore the normal reading frame in mRNA. Frameshift mutations arise in two ways, (1) by error during DNA repair or replication, and (2) by acridine dyes.

Nonsense and missense mutations

After frameshift mutations, three types of codons are produced, (1) sense codons, (2) missense codons, and (3) non-sense condons. The sense codons are normal codons which are read in the same way as before frameshift mutations. Missense mutations have missense codons which code for different amino acid. It usually result in replacement of single amino acid in the polypeptide chain. Missense mutations are more frequent than nonsense mutations. Nonsense mutations are those having nonsense codons which do not code for any amino acid. Nonsense mutations result in premature termination of polypeptide chains and hence are also called **chain terminating mutations**.

Mutagens

Mutagens refer to physical or chemical agents which greatly enhance the frequency of mutations. Various radiations and chemicals are used as mutagens. A brief description of various physical and chemical mutagens is presented below :

Physical mutagens

Physical mutagens include various types of radiations, X-rays, gamma rays, alpha particles, beta particles, fast and thermal (slow) neutrons and ultra violet rays:

X-rays

X-rays were first discovered by Roentgen in 1895 and generated in X-rays machines. They are ionizing and highly penetrating rays. X-rays can break chromosomes and produce all types of mutations in nucleotides, addition, deletion, inversion, transposition, transitions and transversions. These changes are brought out by adding oxygen to deoxyribose, removing amino or hydroxyl group and forming peroxides. X-rays were first used by Muller in 1927 for induction of mutations in Drosophila. In plants, Stadler in 1928 first used X-rays for induction of mutations in barley. Now X-rays are commonly used for induction of mutations in various crop plants. X-rays induce mutations by forming free radicals and ions.

Gamma rays

Gamma rays are identical to X-rays in most of the physical properties and biological effects. But gamma rays have shorter wave length than X-rays and are more penetrating than X-rays. They are generated from radioactive decay of some elements like 14C, 60C, radium etc. Of these, cobalt 60 is commonly used for the production of gamma rays in therapy. Gamma rays cause chromosomal and gene mutations like X-rays by ejecting electrons from the atoms of tissues through which they pass. Now, gamma rays are also widely used for induction of mutations in various crop plants.

Alpha particles

Alpha rays are composed of alpha particles. They are made of two protons and two neutrons and thus have double positive charge. They are densely ionizing, but lesser penetrating than beta rays and neutrons. Alpha particles are emitted by the

isotopes of heavier elements. They have positive charge and hence they are slowed down by negative charge of tissues resulting in low penetrating power. Alpha particles lead to both ionization and excitation resulting in chromosomal mutations.

Beta particles

Beta rays are composed of beta particles. They are sparsely ionizing but more penetrating than alpha rays. Beta particles are generated from radioactive decay of heavier elements such as 3H, 32P, 35S etc. They are negatively charged, therefore, their action is reduced by positive charge of tissues. Beta particles also act by way of ionization and excitation like alpha particles and result in both chromosomal and gene mutations.

UV rays

UV rays are non ionizing radiations, which are produced from mercury vapor lamps or tubes. They are also present in solar radiation. UV rays can penetrate one or two cell layers. Because of low penetrating capacity, they are commonly used for radiation of micro-organisms like bacteria and viruses. In higher organisms, their use is generally limited to irradiation of pollen in plants and eggs in Drosophila. UV rays can also break chromosomes. They have two main chemical effects on pyrimidines.

The first effect is the addition of a water molecule which weakens the H bonding with its purine complement and permits localized separation of DNA strands. The second effect is to join pyrimidines to make a pyrimidine dimmer. This dimerization can produce TT, CC, UU and mixed pyrimidine dimmers like CT. dimerization interferes with DNA and RNA synthesis.

Interstrand dimmers cross link nucleic acid chains, inhibiting strand separation and distribution.

Chemical mutagens

There is a long list of chemicals which are used as mutagens. The chemical mutagens divided into four groups, (1) alkylating agents, (2) base anologues, (3) acridine dyes, and (4) others.

Alkylating agents

This is the most powerful group of mutagens. They induce mutations especially transitions and transversions by adding an alkyl group (either ethyl or methyl) at various positions in DNA. Alkylation produces mutation by changing hydrogen bonding in various ways. The alkylation agents include, ethyl methane sulphonate (MMS), ethylene imines (EI), sulphur mustard, nitrogen mustard, etc. Out of these, the first three are no common use. Since the effect of alkylating agents resemble those of ionizing radiations, they are also known as radiomimetic chemicals. Alkylating agents can cause various large and small deformations of base structure resulting in base pair transitions and transversions. Transversions can occur either because a purine has been so reduced in size that it can accept another purine for its complement, or because a pyrimidine has been so increased in size that it can accept another pyrimidine for its complement. In both cases, diameter of the mutant base pair is close to that of a normal base pair.

Base Analogues

Base analogues refer to chemical compounds which are very similar to DNA bases. Such chemicals sometimes are incorporated in DNA in place of normal base during replication.

Thus, they can cause mutation by wrong base pairing. An incorrect base pairing results in transitions or transversions after DNA replication. The most commonly used base analogues are 5 bromo uracil (5BU) and 2 amino purine (2AP).

5 bromo uracil is similar to thymine, but it has bromine at the C5 position, whereas thymine has CH3 group at C5 position.. The change or shift of hydrogen atoms from one position to another either in a purine or in a pyrimidine base is known as tautomeric shift or (tautomerization).

Acridine Dyes

Acridine dyes are very effective mutagens. Acridine dyes include, proflavin, acridine orange, acridine yellow, acriflavin and ethidium bromide. Out of these, proflavin and acriflavin are in common use for induction of mutation. Acridine dyes get inserted between two base pairs of DNA and lead to addition or deletion of single or few base pairs when DNA replicates. Thus, they cause frameshift mutations and for this reason acridine dyes are also known as frameshift mutagens.

Other mutagens

Other important chemical mutagens are nitrous acid and hydroxylamine. Nitrous acid is a powerful mutagen which reacts with C6 amino groups of cytosine and adenine. It replaces the amino group with oxygen (+ to – H bond). As a result, cytosine acts like thymine and adenine like guanine. Thus, transversions from GC AT and AT GC are induced. Hydroxylamine is a very useful mutagen because it appears to be very specific and produces only one kind of change, namely, the GC AT transition. All the chemical mutagens expect base analogues are known as **DNA modifiers**.

Summary

Mutation is defined as variety of different types of heritable change. New mutations arise in somatic cells or in the germline. Mutations are physiologic that drives evolution (polymorphism), and pathogenic.

Polymorphism

It is the coexistence of several genetically distinctly different types in a population (groups of animals of one species). Examples: different blood groups in humans, different colour forms in some butterflies.

Effects of mutations.

(1) phenotypic abnormality (2)increased susceptibility to disease. (3) mutation in a somatic cell may cause death for that cell (apoptosis) (4) induces inappropriate continuation of cell division, causing cancer.

Origin of mutations

Most mutations arise as copying errors during DNA replication because DNA polymerases are error-prone.

Sources of mutations

Spontaneous: The greatest source of mutations is from spontaneous errors in DNA replication and repair.

Induced: Mutations can be induced in our DNA by exposure to a variety of mutagens occurring in our external environment or to mutagens

Large-scale mutations:

They are chromosome abnormalities which involve loss or gain of chromosomes or breakage and rejoining of chromatids and lead to phenotype abnormality e.g. Dawn syndrome.

Small-scale mutations

They involve a single DNA sequence (simple mutations) or involve exchanges between two allelic or nonallelic sequences (sequence repeat mutation). Three classes of small-scale mutation:

- **Base substitutions:** It involves replacement of usually a single base. **Transitions** are substitutions of a pyrimidine by a pyrimidine, or of a purine by a purine. **Transversions** are substitutions of a pyrimidine by a purine or of a purine by a pyrimidine

Base positions in codons can be grouped into three classes:

Non-degenerate sites: They include the first base position of all but eight codons, the second base position of all codons and the third base position of two codons, **Fourfold degenerate sites** are found at the third base position of several codons. **Twofold degenerate sites** are often found at the third base positions of codons, but also at the first base position in eight codons.

- **Deletions**: One or more nucleotides are eliminated from a sequence.
- **Insertions**: One or more nucleotides are inserted into a sequence.

Mutations in coding and noncoding DNA

Coding and noncoding DNA are about equally susceptible to mutation.

1. **Silent (synonymous) mutations** do not change the sequence of the gene product.
2. **Non-synonymous mutations** result in an altered sequence in a polypeptide or functional RNA. According on their effect:
    a. **Deleterious effect.** can result in disease or death.
    b. **Beneficial effect** improved gene function
    c. **No effect** those with no effect.

Hot spots of pathogenic mutations

Pathogenic mutations can occur at three types of DNA sequence inside the gene locus.

1. The coding sequence of a gene due to nucleotide substitution at first and second base positions of codons (nonsynonymous).
2. Intragenic noncoding sequences necessary for correct expression of the gene, e.g. intronic elements.
3. Regulatory sequences outside exons e.g. promoter elements.

Factors govern mutations expression

1. The location of the mutation within the gene.
2. The presence of a single normal allele with mutation of the other.
3. The degree of mutant expression and other gene products.
4. The proportion and nature of cells in which the mutant gene is present.
5. The parental origin of the mutation.

Gene conversion

This a nonreciprocal transfer of sequence information between a pair of nonallelic DNA sequences (interlocus gene conversion) or allelic sequences

(interallelic gene conversion). The donor, remains unchanged. while the acceptor, is changed .

Sequence repeats mutations

It involves sequence exchange between allelic or non-allelic sequences, often involving repeated sequences.

Genetic mechanisms for VNTR polymorphism

Slipped strand mispairing at short tandem repeats (microsatellites) potential of short tandem repeats is considerable

Unequal crossover: Homologous recombination at meiosis to generate new recombinant strands results in a deletion on one of the participating chromatids and an insertion on the other.

Unequal sister chromatid exchange involving breakage of individual sister chromatIds and rejoining fragments in mitosis

Mitochondrial hotspots

The frequency of 'mitochondrial disorders' is high and the mitochondrial genome can be considered to be a mutation hotspot because (1) the mitochondrial genome has a much higher percentage of coding DNA, (2) it is much more prone to nucleotide change than the nuclear genome (mtDNA instability) and (3) mutations are fixed in the mitochondrial genomes 10 times greater than nuclear genome.

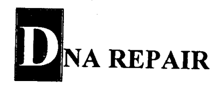
Mutations of mRNA Mutations which alter mRNA sequences can have different consequences:

(1)Failure of splicing causing intron retention (frameshift). (2)Exon skipping. (3)Introduction a premature termination codon .(4) Unstable mRNA.

Frameshift mutations: The mutation which arise due to deletion or additions of nucleotides in mRNA. The reading frame in such case is shifted from the point of addition or deletion onwards.

Nonsense and missense mutations: After frameshift mutations, three types of codons are produced, (1) sense codons, normal codons which are read in the same way as before frameshift mutations (2) missense codons, code for different amino acid and (3) non-sense condons do not code for any amino acid. Nonsense mutations result in premature termination of polypeptide chains and hence are also called **chain terminating mutations**.

CHAPTER 5



DNA REPAIR

Introduction

The DNA sequence can be changed as the result of copying errors introduced by DNA polymerases during replication and by environmental agents such as mutagenic chemicals and certain types of radiation. If DNA sequence changes, whatever their cause, are left uncorrected, both growing and nongrowing somatic cells might accumulate so many mutations that they could no longer function. In addition, the DNA in germ cells might incur too many mutations for viable offspring to be formed. Thus the correction of DNA sequence errors in all types of cells is important for survival.

The relevance of DNA damage and repair to the generation of cancer (carcinogenesis) became evident when it was recognized that all agents that cause cancer (carcinogens) also cause a change in the DNA sequence and thus are mutagens. All the effects of carcinogenic chemicals on tumor production can be accounted for by the DNA damage that they cause and by the errors introduced into DNA during the cells' efforts to repair this damage.

DNA damages:

1. Purine bases are lost by spontaneous fission of the base-sugar link.

2. Cytosines, and occasionally adenines, spontaneously deaminate to produce uracil and hypoxanthine respectively.
3. Many chemicals, for example alkylating agents, form adducts with DNA bases.
4. Reactive oxygen species in the cell attack purine and pyrimidine rings.
5. Ultraviolet light causes adjacent thymines to form a stable chemical dimmer.
6. Mistakes in DNA replication result in incorporation of a mismatched base.
7. Ionizing radiation causes single- or double-strand breaks.
8. Mistakes in replication or recombination leave strand breaks in DNA.

All these lesions must be repaired if the cell is wish to survive. The importance of effective DNA repair systems is highlighted by the severe diseases affecting people with deficient repair systems. To cope with all these forms of damage, cells must be capable of several different types of DNA repair. DNA repair seldom involves simply undoing the change that caused the damage. Almost always a stretch of DNA containing the damaged nucleotide(s) is excised and the gap filled by resynthesis. There are at least six main types of DNA repair mechanisms in human cells:

1- Proofreading by DNA Polymerase

Because the specificity of nucleotide addition by DNA polymerases is determined by Watson-Crick base pairing, a

wrong base (e.g., A instead of G) occasionally is inserted during DNA synthesis. Indeed, the a subunit of E. coli DNA polymerase III introduces about 1 incorrect base in 10^4 internucleotide linkages during replication.

This increased accuracy is due to the proofreading function of DNA polymerases. All DNA. polymerases have a similar three-dimensional structure, which resembles a half-opened right hand. The "fingers" bind the single-stranded segment of the template DNA strand, and the polymerase catalytic activity (Pol) lies in the junction between the fingers and palm. So long as the correct nucleotides are added to the 3′ end of the growing strand, the 3′ end remains in the polymerase site. Incorporation of an incorrect base at the 3′ end causes a melting of the end of the duplex. As a result, the polymerase pauses and the 3′ end of the growing strand is transferred to the 3′ → 5′ exonuclease site (Exo) about 3 nm away, where the mispaired base and probably other bases are removed. Subsequently, the 3′ end flips back into the polymerase site and elongation resumes **(Fig.24)**. It seems likely that this function is indispensable for all cells to avoid excessive genetic damage. Mutations in the gene encoding the e subunit of DNA polymerase III inactivate the proofreading function and lead to a thousand fold increase in the rate of spontaneous mutations.

In addition to the proofreading activity of DNA polymerases that can correct miscopied bases during replication, cells have evolved mechanisms for repairing DNA damaged by chemicals or radiation. Complex organisms with large genomes and relatively long generation times contain many cells that divide very slowly or not at all (e.g., liver and brain cells). Such cells must use the information in their DNA for weeks, months, or even years, greatly increasing their chances for sustaining

damage to their DNA. If repair processes were 100 percent effective, chemicals and radiation would pose no threat to cellular DNA. Unfortunately, repair of lesions caused by some environmental agents is relatively inefficient, and such lesions can lead to mutations that ultimately cause cancer

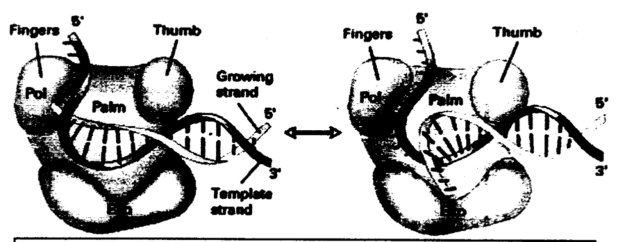


Figure 24. Schematic model of the proofreading function of DNA polymerases.

2-Direct repair

A specific enzyme is able to dealkylate O^6-alkyl guanine directly. In bacteria thymine dimers can be removed in a photoreactivation reaction that depends on visible light and an enzyme, photolyase. Mammals possess enzymes related to photolyase,

3- Base excision repair (BER)

The cell uses glycosidase enzymes to remove abnormal bases. An endonuclease, AP endonuclease, cuts the sugar-phosphate backbone at the position of the missing base. A few nucleotides of the DNA strand are stripped back by exonucleases, the gap is

filled by resynthesis, and the remaining nick is sealed by DNA ligase III. The same process is used to repair spontaneous depurination. Interestingly, no human diseases caused by defective BER are known. May be any such defect would be lethal, since BER corrects much the commonest type of DNA damage.

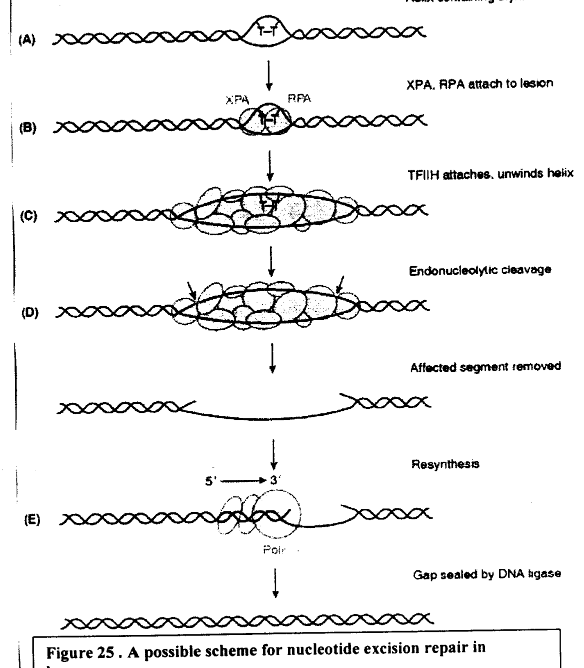


Figure 25 . A possible scheme for nucleotide excision repair in humans.

4- Nucleotide excision repair (NER

This system removes thymine dimers and large chemical adducts. Defects in NER cause the autosomal recessive disease xeroderma pigmentosum. Seven complementation groups, XPA (xeroderma pigmentosum A) -XPG, have been defined by cell fusion studies. XP patients are exceedingly sensitive to UV light. Sun-exposed skin develops thousands of freckles, many of which progress to skin cancer **(Fig. 25)**.

Steps essential for NER are demonstrated as following (1) protein recognizes damaged DNA and binds to it, directly or by binding to RPA, a single-strand binding protein. (2) The DNA-XPA-RPA complex recruits the TFIIH transcription factor (multiprotein complex that includes the XPB and XPD proteins). (3) These are helicases that open a single-stranded bubble in the DNA, about 30 nucleotides long. (4) Two cuts are made in the sugar-phosphate backbone of the damaged strand. XPF + ERCC1 cut at the 5′ end, and XPG cuts the 3′ end. (5) DNA polymerase together with replication factor C and the DPE2 subunit synthesize DNA to fill the gap. (6) DNA ligase seals the gap. Over 30 proteins are involved in mammalian nucleotide excision repair.

5- Post-replication repair

It is required to correct double-strand breaks. The usual mechanism is a gene conversion-like process (recombinational repair), where a single strand from the homologous chromosome invades the damaged DNA helix. Alternatively, broken ends are rejoined regardless of their sequence, a desperate measure that is likely to cause mutations. The eukaryotic machinery for

recombination repair is less well defined than the excision repair systems. Human genes involved in this pathway include NBS (mutated in Nijmegen breakage syndrome, MIM), BLM (mutated in Bloom syndrome, MIM) and the BRCA2 and maybe BRCA1 breast cancer susceptibility genes(**Fig. 26**).

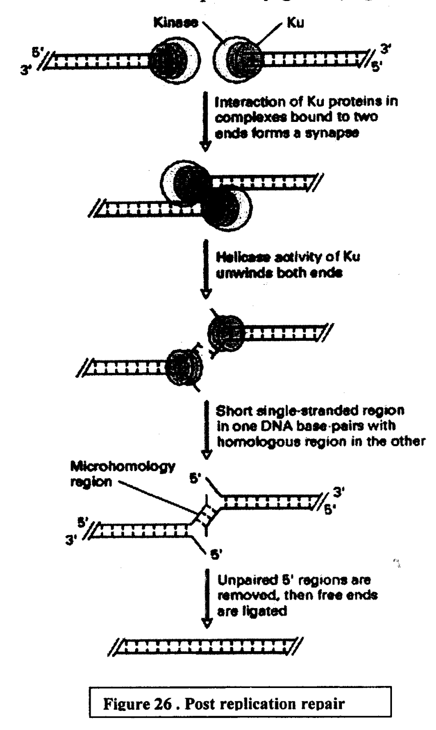


Figure 26 . Post replication repair

The post replication repair needs the following steps (1)A complex of two proteins, Ku and DNA-dependent protein kinase, binds to the ends of a double-strand breaks. (2) After formation of a synapse in which the broken ends overlap, Ku unwinds the ends, (3) This mechanism yields a region of microhomology (base pairs are more or less complimentary). (4) The unpaired single-stranded 5′ ends are removed by mechanisms that are not well understood, and (6) The two double-stranded molecules ligated together. As a result, the double-strand break is repaired, but several base pairs at the site of the break are removed.

6- Mismatch repair

Corrects mismatched base pairs caused by mistakes in DNA replication. Cells deficient in mismatch repair have mutation rates 100 - 1000 times higher than normal, with a particular tendency to replication slippage in homopolymeric runs. In humans the mechanism involves at least five proteins and defects cause.

As shown in **Fig. 27,** a replication error introduces a mismatch occurs through (A) The MutS (mutation) protein binds to mismatched base pairs (B) In an ATP-dependent reaction, a MutS-MutL-MutH complex is formed which probably brings any GATC sequence located within 1 kb either side of the mismatch into a loop (C) MutH makes a single-strand cut 5′ to the GATC sequence (D) The methylation system methylates Adenin in GATC, but in newly synthesized DNA only the template strand is methylated. MutH specifically cuts the unmethylated (newly synthesized) strand (E) Exonucleases, DNA polymerase and DNA ligase then strip back and repair the DNA.

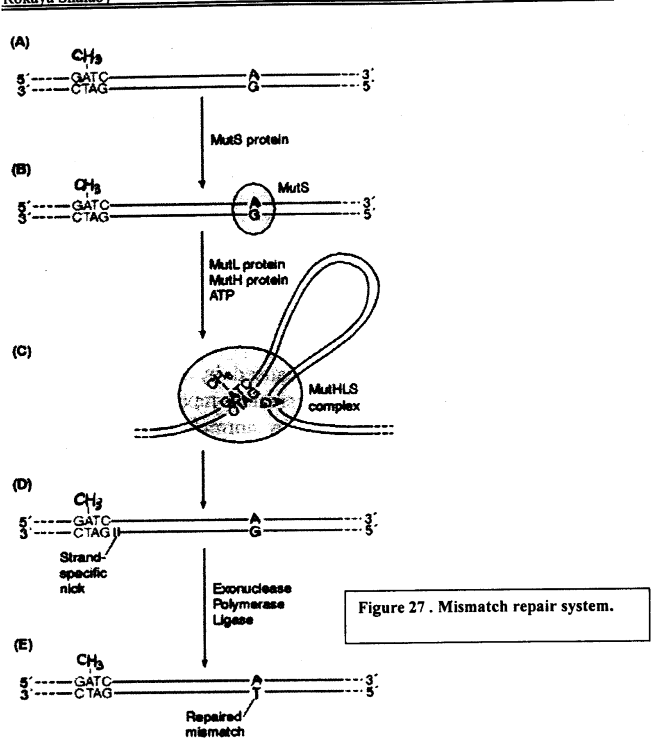


Figure 27 . Mismatch repair system.

All these six systems, except for direct repair, require exo- and endonucleases, helicases, polymerases and ligases, usually acting in multiprotein complexes that have some components in

common. Sorting out the individual pathways has been greatly aided by the very strong conservation of repair mechanisms across the whole spectrum of life. Generally eukaryotes have multiple proteins corresponding to each single protein in E. coli, so that, for example, nucleotide excision repair requires six proteins in E. coli but at least 30 in mammals.

As well as sharing components with each other, many repair systems share components with the machinery for DNA replication, transcription and recombination. DNA polymerases and ligase are required for both DNA replication and resynthesis after excision of a defect. The recombination machinery is involved in double-strand break repair. The link with transcription is particularly intriguing. The general transcription factor TFIIH is a multiprotein complex that includes the XPB and XPD proteins. TFIIH exists in two forms. One form is concerned with general transcription and the other with repair, probably specifically repair of transcriptionally active DNA. This system is deficient in two rare diseases, Cockayne syndrome and trichothiodystrophy. Clinically and in cell biology, CS and TTD both overlap XP, and in some cases the same genes are responsible, but CS and TTD patients have developmental defects that presumably reflect defective transcription, and they do not have the cancer susceptibility of XP patients.

Many human diseases that involve hypersensitivity to DNA-damaging agents, or a high level of cellular DNA damage, are not caused by defects in the DNA repair systems themselves, but by a defective cellular response to DNA damage. Normal cells react to DNA damage by stalling progress through the cell cycle at a checkpoint until the damage has been repaired, or

triggering apoptosis if the damage is irrepairable. Part of the machinery for doing this involves the ATM protein. Briefly, it senses DNA damage and relays the signal to the p53 protein, the 'guardian of the genome'. People with no functional ATM have ataxia telangiectasia. Their cells are hypersensitive to radiation, and they have chromosomal instability and a high risk of malignancy, but the DNA repair machinery itself is intact. Fanconi anemia is another heterogeneous group of diseases (at least five complementation groups) marked by defective responses to DNA damage, without specific defects in DNA repair.

Summary

The DNA sequence can be changed as the result of copying errors introduced by DNA polymerases during replication and by environmental agents such as mutagenic chemicals and certain types of radiation.

DNA damages:

1. Purine bases are lost by spontaneous fission of the base-sugar link.
2. Cytosines, and occasionally adenines, spontaneously deaminate to produce uracil and hypoxanthine respectively.
3. Many chemicals, for example alkylating agents, form adducts with DNA bases.
4. Reactive oxygen species in the cell attack purine and pyrimidine rings.
5. Ultraviolet light causes adjacent thymines to form a stable chemical dimmer.
6. Mistakes in DNA replication result in incorporation of a mismatched base.
7. Ionizing radiation causes single- or double-strand breaks.
8. Mistakes in replication or recombination leave strand breaks in DNA.

All these lesions must be repaired if the cell is wish to survive.

1- Proofreading by DNA Polymerase

This function is indispensable for all cells to avoid excessive genetic damage. Mutations in the gene encoding the e subunit of DNA polymerase III inactivate the proofreading function and lead to a thousand fold increase in the rate of spontaneous mutations.

DNA polymerases can correct miscopied bases during replication,.

2-Direct repair

In bacteria thymine dimers can be removed in a photoreactivation reaction that depends on visible light and an enzyme, photolyase. Mammals possess enzymes related to photolyase,

3- Base excision repair (BER)

The cell uses glycosidase enzymes to remove abnormal bases. An endonuclease, , cuts the sugar-phosphate backbone at the position of the missing base. A few nucleotides of the DNA strand are stripped back by exonucleases, the gap is filled by resynthesis, and the remaining nick is sealed by DNA ligase III.

4- Nucleotide excision repair (NER

This system removes thymine dimers and large chemical adducts. Seven complementation groups, XPA (xeroderma pigmentosum A) -XPG, have been defined by cell fusion studies. Defects in NER cause the autosomal recessive disease xeroderma pigmentosum.

5- Post-replication repair

A complex of two proteins, binds to the ends of a double-strand breaks and formation of a synapse in which the broken ends overlap. This mechanism yields a region of microhomology (base pairs are more or less complimentary). The unpaired single-stranded 5′ ends are removed by mechanisms that are not well understood, and the two double-stranded molecules ligated together. As a result, the double-strand break is repaired, but several base pairs at the site of the break are removed.

6- Mismatch repair

Corrects mismatched base pairs caused by mistakes in DNA replication.

Many human diseases involve hypersensitivity to DNA-damaging agents, and are not caused by defects in the DNA repair systems. Normal cells react to DNA damage by stalling progress through the cell cycle at a checkpoint until the damage has been repaired, or triggering apoptosis if the damage is irrepairable.

CHAPTER 6

CELL DIVISION

Introduction

Cell is the basic unit of structure and function in all organisms. The process of reproduction of formation of new cells from the pre-existing cells is referred to as cell division. The cell which undergoes division is known as mother cell and the new cells which are formed by the process of cell division are termed as daughter cells. As a person develops from an embryo through fetus and infant to an adult, cell divisions are needed to generate the large numbers of required cells. Additionally, many cells have a limited lifespan, so there is a continuous requirement to generate new cells in the adult. All these cell divisions occur by mitosis. Mitosis is the normal process of cell division, from cleavage of the zygote to death of the person. In the lifetime of a human there may be something like 10^{17} mitotic divisions. There are two types of cell division, mitosis and meiosis. They are briefly described below, **(Fig.28)**:

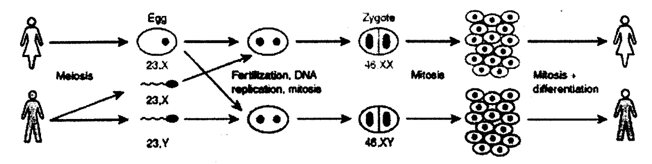


Figure 28. Human life, from a chromosomal viewpoint. The haploid sperm and egg cells originate by meiosis from diploid precursors. In the fertilized egg the sperm and egg chromosomes initially form separate male and female pronuclei. These combine during the first mitosis.

Cell cycle

The period in which one cycle of cell division is completed is called cell cycle. A cell cycle consists of two phases, Interphase and mitosis (M phase) **(Fig.29).**

Mitosis

The term mitosis was coined by Flemming in 1882. mitosis refers to the spindle using nuclear division which produces two identical daughter nuclei from parent nucleus. Since mitosis occurs in somatic cells it is also called **somatic cell division**. The important feature of mitosis are briefly described below:

1. Mitosis leads to production of two daughter cells from a mother cell in each cycle of cell division. In other words nucleus divides once in each cell cycle.
2. The daughter cells are similar to mother cell in shape, size and chromosome complement. Since the chromosome number in the daughter cells is the same as in the mother cell this is also known as homotypic or educational division.
3. Segregation and recombination do not place during mitosis.

The M phase of the cell cycle (mitosis) consists of the four stages of nuclear division (prophase, metaphase, anaphase and telophase), and cell division (cytokinesis), which overlaps the final stages of mitosis.

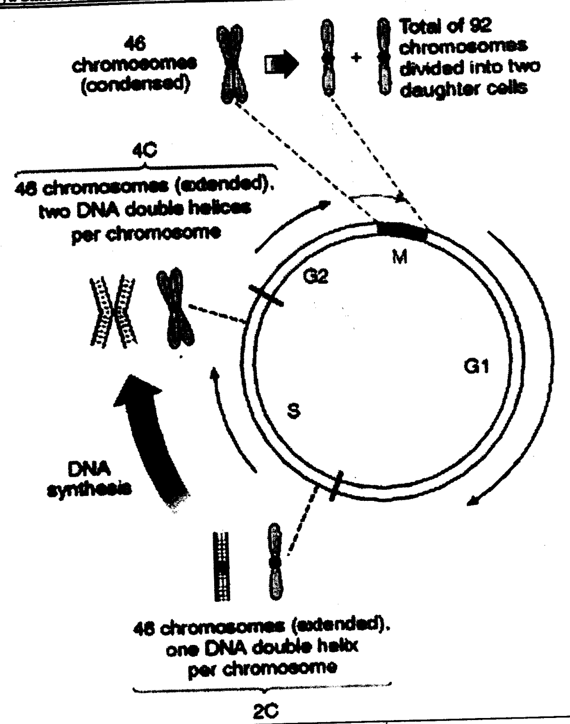


Figure 29. Human chromosomal DNA content during the cell cycle.
Interphase comprises $G_1 + S + G_2$.
Chromosomes contain one DNA double helix from anaphase until S phase.
From S phase until the end of metaphase, the chromosome consists of two chromatids each containing a DNA duplex.
The DNA content of a diploid cell before S phase is 2C (twice the DNA content of a haploid cell), while between S phase and mitosis it is 4C.

Prophase

Prophase starts immediately after G2 stage of interphase. Chromosomes are thin and uncoiled in the early prophase, but become shortened, coiled and more distinct during mid prophase. In the late prophase, chromosomes appear more conspicuous, short and thick and longitudinally double. The two chromatids of each chromosomes held at centromere are visible under light microscope. The nuclear membrane and nucleolus disappear at the end of prophase.

Metaphase

In preparation for cell division, the previously highly extended chromosomes contract and condense so that, at metaphase, they are readily visible under the microscope. Even though the DNA was replicated some time previously, it is only at prometaphase the individual chromosomes can be seen to comprise two sister chromatids, attached at the centromere.

The mitotic spindle is formed from tubulin-based microtubules and microtubule-associated proteins. **Polar fibers, (Fig.30)**which extend from the two poles of the spindle towards the equator, develop at prophase while the nuclear membrane is still intact. Kinetochore fibers do not develop until prometaphase.

These fibers attach to the **kinetochore, (Fig.31)** (a large multiprotein structure attached to the centromere of each chromatid), and extend in the direction of the spindle poles. The interaction between the different spindle fibers pulls the chromosomes towards the center, and by metaphase each

chromosome is independently aligned on the equatorial plane (metaphase plate).

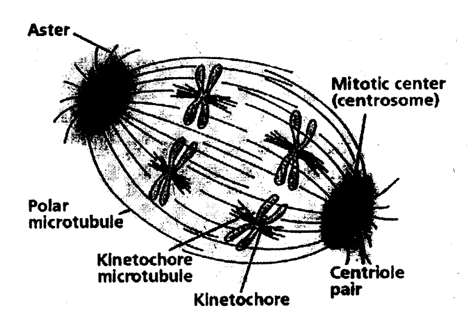


Figure30. Mitosis: At metaphase, paternal and maternal homologs of each chromosome pair are independently aligned at the metaphase plate, and not associated with each other. Microtubules attached to the kinetochores link chromosomes to each of the poles. Other spindle microtubules include astral microtubules that radiate from each pole, and polar microtubules that form attachments linking the two poles.

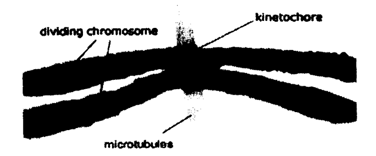


Figure 31. Kinetochore formation in metaphase chromosome attached to spindle fibers

Anaphase

The spindle fibers pull the separated sister chromatids of each chromosome to opposite poles **(Fig.32).** Paternal and maternal homologs do not associate at all during mitosis. Following centromere division at anaphase, The DNA of the two sister chromatids is identical, barring any errors in DNA replication. Thus the effect of mitosis is to generate daughter cells that contain precisely the same DNA sequences.

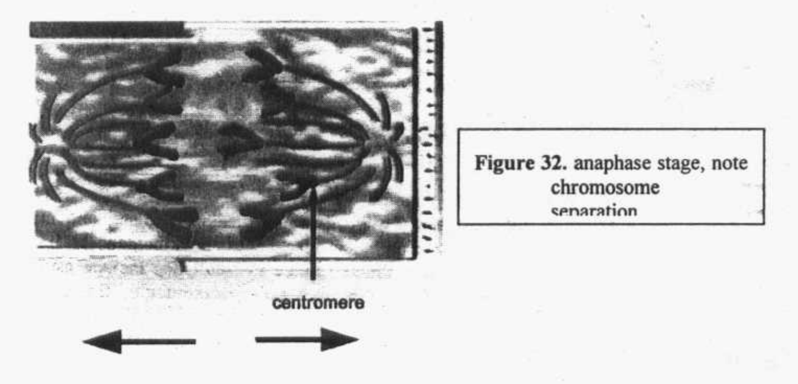


Figure 32. anaphase stage, note chromosome separation

Telophase

When chromosomes reach the pole, the last stage , Telophase begins. The spindle tubes disintegrate and a new nuclear envelope is formed at each pole covering the chromosomes. The nucleoli also appear at each pole **(Fig.33).** Chromosomes again become thinner and longer by uncoiling and unfolding and look like a single thread under light microscope. Then the nucleus enters interphase. Among all the four phases of mitosis, prophase takes longest duration.

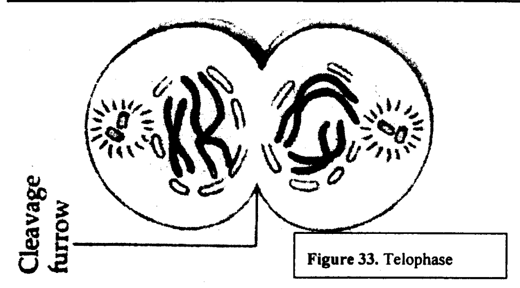


Figure 33. Telophase

Cytokinesis

The division of nucleus is known as **karyokinesis**. It is followed by division of cytoplasm, which is known as cytokinesis. The division of cytoplasm into tow daughter cells may take place in animals, when the separation of cytoplasm starts by furrowing of plasma lemma in the equatorial region. This results in division of cytoplasm into two daughter cells.

Genetic control of mitosis

The mitotic cell division is under genetic control in all eukaryotes under normal conditions. Important evidences in this connection are given below:

1. In eukaryotes, under similar environmental conditions, cells do not divide continuously. They sometimes divide and sometimes do not divide. This indicates that on and off mechanism must be genetically controlled.
2. Some cells divide more frequently than others. For example, somatic cell of an embryo can undergo a larger number of mitotic divisions than a somatic cell of an adult. This

difference in the mitotic division is also due to genetic control at two different stages. Highly differentiated cells like nerve and cardiac muscle cells lose its capacity to divide. The genetic control is lost in somatic cells which become cancerous. Hence such cells divide mitotically continuously without any control resulting in abnormal tissue growth.

3. The replication of each chromosome precisely into two chromatids during S stage of interphase again provides strong evidence that mitosis is genetically controlled

Meiosis

Primordial germ cells migrate into the embryonic gonad and engage in repeated rounds of mitosis (many more in males than in females) to form oogonia in females and spermatogonia in males **(Fig.33).** Further growth and differentiation produces primary oocytes in the ovary and primary spermatocytes in the testis. These specialized diploid cells can undergo meiosis. Meiosis involves two successive cell divisions but only one round of DNA replication, so the products are haploid cells. In males, the product of one primary spermatocyte is four spermatozoa. In females, the cytoplasm divides unequally at each stage: the products of meiosis I (the first meiotic division) are a large secondary oocyte and a small cell (polar body). The secondary oocyte then gives rise to the large mature egg cell and a second polar body. The important feature of mitosis are briefly presented below:

1. Meiosis leads to production of four daughter cells from a mother cell in each cycle of cell division. In other words nucleus divides twice in each cell cycle
2. The daughter cells are similar to mother cell in shape and, size but different in chromosome complement.
3. The daughter cells have haploid chromosome number. Meiosis occurs in reproductive organs like ovaries and testis.
4. The complete process of meiosis consists of two types of divisions. The first division results in reduction of chromosome number to half and is called reductional division. The second division is like mitotic division
5. Meiosis results in segregation of chromosome and genes and their independent assortment. Crossing over recombination also occur during meiosis.

Mitosis involves a single turn of the cell cycle. The DNA is replicated in S phase and the two copies are divided exactly equally between the daughter cells in M phase. Meiosis is also preceded by one round of DNA synthesis, but then there are two cell divisions without intervening DNA synthesis, so that the products end up haploid. The second division of meiosis is identical to mitosis, but the first division has important differences whose purpose is to generate genetic diversity between the daughter cells. This is done by two mechanisms, independent **assortment** of paternal and maternal homologs, and **recombination**.

Meiosis I

The first meiotic division results in reduction of chromosome number in each new cell to just half of the mother cell, therefore, it is referred to as reductional division. The first meiotic division consistes of four different phases, (1) prophase, (2) , metaphase, (3) anaphase and (4) telophase.

First prophase

This phase starts after interphase and is of maximum duration. This consists of five sub stages, leptotene, zygotene, pachytene, diplotene and diakinesis. Important features of these sub stages are briefly discussed below:

Leptotene

1. Chromosomes look like thin thread under light microscope. They are inter-woven like a losse ball of wool.
2. Chromosomes are scattered throughout the nucleus in a random manner.
3. In some cases, chromomeres are visible on the chromosomes in the form of condensed regions.
4. RNA and protein syntheses also take place.

Zygotene

1. Homologous chromosomes being to pair.
2. Chromosomes become shorter and thicker .
3. The synthesis of remaining 0.3% DNA which has not taken place during S phase also occurs during this stage.
4. Synaptonemal complex also develops during this stage **(Fig.34).**

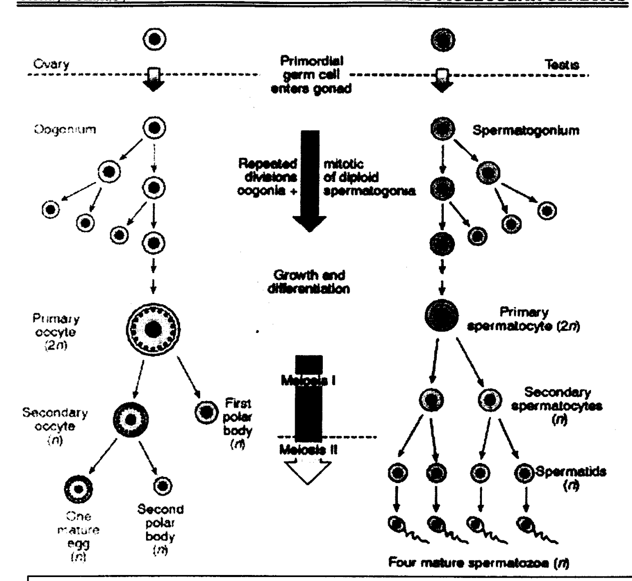


Figure 34. Development of the germ line.

The germ line produces primary oocytes and primary spermatocytes.

Meiosis involves two cell divisions but only one DNA replication, so the products are haploid.

In humans, primary oocytes enter meiosis I during fetal life but then arrest at the prophase stage right through to puberty or later.

After puberty, one oocyte a month completes meiosis.

Sperm are produced continuously from puberty onwards.

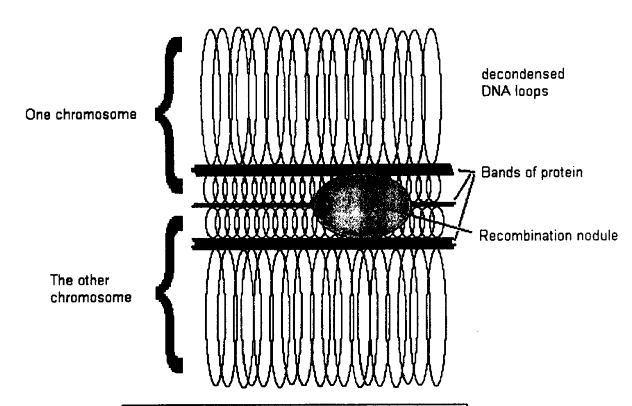


Figure 35. Diagram for synaptonemal complex

Pachytene

1. Chromosomes look like bivalents. Each bivalent has two chromatids. Thus each pair has four chromatids generally known as tetrads
2. The chromosome number looks like haploid number.
3. Nucleolus is present and attached to a chromosome.
4. Formation of chiasma an crossing over take place during pachytene stage.**(Fig.35)**

Diplotene

1. Separation of homologous chromosomes begins. It starts at centromere and moves towards the end.
2. The separating chromosomes are attached at some points. These points are called chiasmata. These chiasmata are terminalized towards the end of diplotene.

3. Chromosomes are further condensed and become still shorter and thicker.
4. Nucleolus decreases in size.

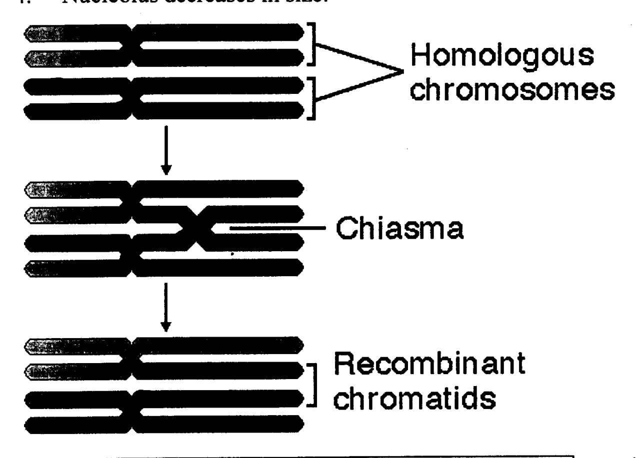


Figure 36. Chiasma formation between two homologous chromosomes

Diakinesis

1. This stage begins after complete terminilization of chiasmata.
2. Chromosomes are further condensed.
3. Bivalents are distributed throughout the cell.
4. Nucleolus and nuclear membrane disappear towards the end of diakinesis.

Segregation

During meiosis I the maternal and paternal homologs of each chromosome pair form a bivalent by pairing together (synapsis). Each chromosome consists of two sister chromatids following DNA replication, so that the bivalent is a four-stranded structure at the metaphase plate. Spindle fibers then pull one complete chromosome (two chromatids) to either pole. However, for each of the 23 homologous pairs, the choice of which homolog enters which daughter cell is independent. This allows 2^{23} or about 8.4×10^6 possible combinations of parental chromosomes to be produced by one person.

Recombination

- During prophase of meiosis I the synapsed homologs within each bivalent exchange segments in a random way.
- At the zygotene stage, each pair of homologs begins to form a synaptonemal complex consisting of the two chromosomes in close apposition, separated by a long linear protein core.
- Completion of this complex marks the start of the pachytene stage, which is when recombination (or crossover) occurs.
- Crossing-over involves physical breakage of the double helix in one paternal and one maternal chromatid, and joining of maternal and paternal ends.
- Overall, the combination of recombination between homologs in prophase I plus independent assortment of homologs at anaphase I ensures that a single individual can

produce an almost unlimited number of genetically different gametes.

- The mechanism allowing alignment of the homologs is not understood. However, it is thought that such close apposition is required for recombination. Recombination nodules, very large multiprotein assemblies located at intervals on the synaptonemal complex, are thought to mediate the recombination events.
- The two homologs can be seen to be physically connected at specific points. Each such connection is described as a **chiasma** (plural chiasmata) **(Fig.36)** and marks a crossover point. There are an average of 55 chiasmata in a male meiotic cell, and may be 50% more in female meiosis. The genetic consequences of crossing over are considered in.
- In addition to their role in recombination, chiasmata are thought to be essential for correct chromosome segregation at meiosis I. By holding the maternal and paternal homologs of each chromosome pair together on the spindle until anaphase I, they have a role that is analogous to that of the centromeres in mitosis and meiosis II.
- There is genetic evidence that children with wrong numbers of chromosomes are often the product of gametes where a bivalent lacked crossovers.

Meiosis II

It appears identical to mitosis, except that there are only 23 chromosomes instead of 46. Each chromosome consists of two chromatids, and these are separated in anaphase II. However,

there is one difference. The sister chromatids of a mitotic chromosome are identical, being copies of each other. The two chromatids of a chromosome in meiosis II may be genetically different as a result of crossovers in meiosis I

X-Y pairing

In female meiosis, each chromosome has a fully homologous partner, and the two Xs synapse and crossing over like any other pair of homologs (autosomes). In male meiosis there is a problem. The human X and Y sex chromosomes are very different. Nevertheless, they pair in prophase I in males, thus ensuring that at anaphase I each daughter cell receives one sex chromosome, either the X or the Y. X-Y pairing is end-to-end rather than along the whole length, and it is made possible by a 2.6 Mb region of homology between the X and Y chromosomes at the tips of their short arms. Pairing is sustained by an obligatory crossover in this region. Genes in the pairing segment have some interesting properties:

1. They are present as homologous copies on the X and Y chromosomes.
2. They are not subject to X-inactivation (as expected since each sex has two copies).
3. Because of the crossing over, alleles at these loci do not show the normal X-linked or Y-linked patterns of inheritance, but segregate like autosomal alleles.

Pseudoautosomal region

Because of this behavior, this region is known as the major pseudoautosomal region. A second smaller pseudoautosomal region of 320 kb is located at the tips of the long arms of both chromosomes, but pairing and crossing-over in this minor pseudoautosomal region is not an obligatory feature of male meiosis.

Genetic control of meiosis

Various features of meiosis are controlled by genes. Some of the features of meiosis which are genetically controlled are described below:

1. **Synapsis and exchange**. synapsis or pairing between homologous chromosomes may depend on the presence of a specific allele. For example, in maize when this allele is absent, synapsis is prevented between all homologous loci. In the absence of synapsis, no exchange occurs between homologous chromosomes and distribution of chromosomes is also irregular during anaphase I. In Drosophila male, crossing over does not occur because the homolgous chromosomes pair only in the heterochromatic region near centromere. Heterochromatin is considred devoid of active genes, hence exchange is prevented.
2. **Centromere Behavior**. The specific behavior of centromere during meiosis is genetically controlled. At metaphase II, centromeres of sister chromatids lie very close together. But when one of them faces one pole of spindle, the other one automatically faces the opposite pole. In Drosophila, in

the presence of a particular allele, sister centromeres separate early at metaphase II and orient to the spindle independently. As a result, both sister centromeres sometimes orient to the same pole and hence are not distributed to daughter nuclei properly. However, this allele has no effect on mitosis or meiosis I.

3. **Spindle shape**. The shape of spindle is governed by a specific allele. The presence of abnormal alleles change the shape of spindle in maize and Drosophila during meiosis but not mitosis. A normal allele causes the meiotic spindle to have convergent ends. With such spindle, all the chromosomes come together in group to the pole and are included in a telophase nucleus. In the presence of abnormal allele, divergent spindles are formed. Such spindles lead to the spread of chromosomes at anaphase in such a way that some are left out of telophase nuclei.
4. **Spindle orientation**. The orientation is usually similar in meiosis I and meiosis II. This leads to formation of four nuclei or cells. When this direction is opposite to that of meiosis I, the result is a cluster of four nuclei or cells.

There are some differences between mitosis and meiosis, which are briefly presented in **Table 4.**

Table 4. Differences between mitosis and meiosis.

Mitosis	Meiosis
1. Consists of one nuclear division	Consists of two nuclear division
2. One cell cycle results in production of two daughter cells	One cell cycle results in the formation of 4 daughter cells
3. The chromosome number of daughter cells is the same as that of mother cell (2n)	Daughter cells contain half the chromosome number of mother cell (n)
4. Daughter cells are identical with mother cell in structure and chromosome composition	Daughter cells are different from mother cell in chromosome number and composition
5. Mitosis occurs in somatic tissues	It occurs in reproductive tissues
6. Total DNA of nucleus replicates during S phase	99.7%DNA replicates During S phase and remaining 0.3% during zygotene stage
7. There is no pairing between homologous chromosomes.	Homologous chromosomes pair during pachytene.
8. Segregation and recombination do not occur	Crossing over takes place during pachytene
9. Chromosomes are in the from of dyad at metaphase	Chromosomes are in the from of tetrad at metaphase
10. At metaphase, centromere of each bivalent divides longitudinally	The centromere dose not divide at metaphase
11. One member of sister chromatids move to opposite pole during anaphase	One member of homologous chromosomes moves to opposite pole during anaphase I
12. Maintains purity due to lack of segregation and recombination	Generates variability due to segregation and recombination.

Summary

Cell is the basic unit of structure and function in all organisms. There are two types of cell division mitosis and meiosis.
Mitosis leads to production of two daughter cells similar in shape, size and chromosome number and structure. It consists of the four stages.
In prophase chromosomes are thin and uncoiled and become shortened, coiled and more distinct during mid prophase. In the late prophase, chromosomes appear more, short and thick
In metaphase chromosomes condense, and become visible under the microscope comprised from two sister chromatids, attached at the centromere attached to the mitotic by Kinetochore fibers
In anaphase the sister chromatids separated by movement of spindle fibers to opposite poles. .
In telophase the spindle tubes disintegrate, A new nuclear envelope is formed at each pole covering the chromosomes. The nucleoli also appear at each pole.
Cytokinesis means division of cytoplasm,.
The mitotic cell division is under genetic control. This proved by(1) cells sometimes divide and sometimes do not divide. (2) Some cells divide more frequently than others. (3) The replication of each chromosome precisely into two chromatids during S stage.
Meiosis
Meiosis involves two successive cell divisions but only one round of DNA replication, it occur in ovary and testis . in male one cell produces 4 sperms in female one cell produce one egg. The daughter cells are different in chromosome structure (recombination and segregation) and number (haploid) .
Meiosis I The first meiotic division results in reduction of chromosome number in each new cell to just half of the mother cell,. it consists of, (1) prophase, (2) , metaphase, (3) anaphase and (4) telophase.
First prophase
This phase starts after interphase and is of maximum duration. This consists of five sub stages, leptotene, zygotene, pachytene,

In leptotene chromosomes are thin thread and homologous chromosomes being to pair. Synaptonemal complex also develops.
In pachytene chromosomes are bivalents. Each bivalent has two chromatids.
Zygotene. Thus each pair has four chromatids. Formation of chiasma an crossing over take place
In diplotene Separation of homologous chromosomes begins. The separating chromosomes are attached at some points. These points are called chiasmata. These chiasmata are terminalized towards the end of diplotene.
diakinesis after complete terminilization of chiasmata. chromosomes are condensed.
Segregation means independency of movement of paternal and maternal chromosomes in gamete cells to daughter cell. This allows 2^{23} or about 8.4×10^6 possible combinations of parental chromosomes to be produced by one person.
Recombination Crossing-over involves physical breakage of the double helix in one paternal and one maternal chromatid, and joining of maternal and paternal ends. The two homologs can be seen to be physically connected at specific points. Each such connection is described as a **chiasma**

Meiosis II It appears identical to mitosis

X-Y pairing X-Y pairing is end-to-end rather than along the whole length. Because of this behavior, this region is known as the major pseudoautosomal region. A second smaller pseudoautosomal region located at the tips of the long arms of both chromosomes,.
Genetic control of meiosis. meiosis is controlled by genes because of (1) ynapsis or pairing between homologous chromosomes may depend on the presence of a specific genes, (2)the specific behavior of centromere, (3) the shape of spindle is governed by a specific allele and (4) orientation(the orientation of spindle is usually similar in meiosis I and meiosis II).

CHAPTER 7

CHROMOSOMES STRUCTURE AND FUNCTIONS

Introduction

Chromosomes as seen under the microscope and illustrated in textbooks are rather misleading. When we look at chromosomes in a dividing cell we see the genome of the cell largely switched off and packed up into neat bundles ready for cell division. The processes of cell division are fascinating in their own right, and errors in packaging or dividing up the genome have major medical consequences.

Interphase chromosome

However, it is important to remember that the switched on, functioning interphase chromosome that controls cellular activities is a much more extended and diffuse structure than the metaphase chromosomes. Importantly, it comprises only a single chromatid and one DNA double helix, not the two-chromatid structure of mitotic chromosomes. As functioning organelles, eukaryotic chromosomes seem to require only three classes of DNA sequence element: 1-centromeres, 2-telomeres and 3-origins of replication **(Fig.37)**.

This simple requirement has been verified by the successful construction of artificial chromosomes in yeast: large foreign DNA fragments behave as autonomous chromosomes

when ligated to short sequences that specify a functional centromere, two telomeres and a replication origin. Recently mammalian artificial chromosomes have been constructed on similar principles.

Centromere

TCACATGAT AGTGTACTA	80 90 bp > 90% (A+T)	TGATTTCCGAA ACTAAAGGCTT
I	II	III

Telomere

Tandem repeats based on the general formula

$(TG)_{1-3}TG_{2-3}C_{2-3}A(CA)_{1-3}$

e.g.

5'...TGTGTGGGTGTGGTGTGTGTGG 3'
3'....ACACACCCACACCACACACACC 5'

Autonomous replicating sequence

Contains an 11-bp core consensus that is AT-rich, plus some imperfect copies of this sequence spanning an approximately 50-bp region of DNA

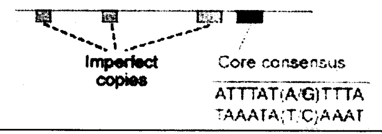


Figure 37. The functional elements of chromosome centromere origin of replication and telomrer

Centromere

Normal chromosomes have a single centromere that is seen under the microscope as the primary constriction, the region at which sister chromatids are joined. The centromere is essential for segregation during cell division. Chromosome fragments that lack a centromere (acentric fragments) do not become attached to the spindle, and so fail to be included in the nuclei of either of the daughter cells.

During late prophase of mitosis, a pair of kinetochores forms at each centromere, one attached to each sister chromatid. Multiple microtubules attach to each kinetochore, linking the centromere of a chromosome and the two spindle poles. At anaphase, the kinetochore microtubules pull the two sister chromatids toward opposite poles of the spindle. Kinetochores play a central role in this process, by controlling assembly and disassembly of the attached microtubules and, through the presence of motor molecules, by ultimately driving chromosome movement.

Specific DNA sequences presumably specify the structure and function of centromeres. In simple eukaryotes, the sequences that specify centromere function are very short. For example, in the yeast Saccharomyces cerevisiae the centromere element (CEN) is about 110 bp long, comprising two highly conserved flanking elements of 9 bp and 11 bp and a central AT-rich segment of about 80-90 bp. The centromeres of such cells are interchangeable - a CEN fragment derived from one yeast chromosome can replace the centromere of another with no apparent consequence. In mammals, centromeres comprise

hundreds of kilobases of repetitive DNA, some nonspecific and some are chromosome-specific.

Centromere has four main functions: (1) Orientation of chromosomes at metaphase, (2) movment of chromosomes at anaphase, (3) formation of chromatid and (4) shape of chromosomes.

Origins of replication

The DNA in most diploid cells normally replicates only once per cell cycle. The initiation of replication is controlled by cis-acting sequences that lie close to the points at which DNA synthesis is initiated. Probably these are sites at which trans-acting proteins bind.

Eukaryotic origins of replication have been most comprehensively studied in yeast, where the presence of a putative replication origin can be tested by a genetic assay. To test the ability of a random fragment of yeast DNA to promote autonomous replication, it is incorporated into a bacterial plasmid together with a yeast gene that is essential for growth of yeast cells. This construct is used to transform a mutant yeast that lacks the essential gene. The transformed cells can only form colonies if the plasmid can replicate in yeast cells. However, the bacterial replication origin in the plasmid does not function in yeast, therefore the few plasmids that transform at high efficiency must possess a sequence within the inserted yeast fragment that confers the ability to replicate extrachromosomally at high

efficiency, that is an autonomously replicating sequence (ARS) element.

ARS elements are thought to derive from authentic origins of replication and, in some cases, this has been confirmed by mapping a specific ARS element to a specific chromosomal location and demonstrating that DNA replication is indeed initiated at this location. ARS elements extend for only about 50 bp and consist of an AT-rich region which contains a conserved core consensus and some imperfect copies of this sequence. In addition, the ARS elements contain a binding site for a transcription factor and a multiprotein complex is known to bind to the origin of replication.

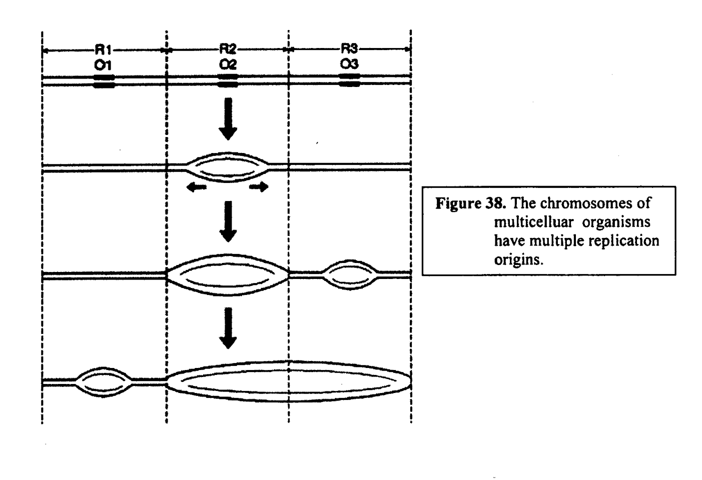


Figure 38. The chromosomes of multicelluar organisms have multiple replication origins.

Mammalian replication origins have been much less well defined because of the absence of a genetic assay. Some initiation sites have been studied, but such studies have not been able to identify a unique origin of replication **(Fig.38)**. This has led to speculation that replication can be initiated at multiple sites over regions which are tens of kilobases long. Mammalian artificial chromosomes seem to work without specific ARS sequences being provided. Computer analysis of regions encompassing several eukaryotic origins of replication, including some human and other mammalian examples, identified a consensus DNA sequence WAWTTDDWWWDHWGWHMAWTT where W = A or T; D = A or G or T; H = A or C or T; and M = A or C

Telomeres

Telomeres are specialized structures, comprising DNA and protein, which cap the ends of eukaryotic chromosomes. Eukaryotic telomeres consist of a long array of tandem repeats. One DNA strand contains TG-rich sequences and terminates in the 3 end; the complementary strand is CA-rich. Unlike centromeres, the sequence of telomeres has been highly conserved in evolution. There is considerable similarity in the simple sequence repeat, for example TTGGGG (Paramecium), TAGGG (Trypanosoma) TTTAGGG (Arabidopsis) and TTAGGG (Homo sapiens) **(Fig.39)**. Telomeres have several likely functions:

1. Maintaining the structural integrity of a chromosome. If a telomere is lost, the resulting chromosome end is unstable. It has a tendency either to fuse with the ends of other broken

chromosomes, to be involved in recombination events or to be degraded. The loop structure of human telomeres means that natural chromosomes have no free DNA end.

2. Ensuring complete replication of the extreme ends of chromosomes. During DNA replication, synthesis of the lagging strand is discontinuous and requires the presence of some DNA ahead of the sequence, which is to be copied serve as the template for an RNA primer. However, at the extreme end of a linear molecule, there no such a template, and a different mechanism is required to solve the problem of replicating the ends of a linear DNA molecule. The problem of replicating the ends of a chromosome has been solved by extending the synthesis of the leading strand using a specialized enzyme, telomerase. This RNA-protein complex carries within its RNA component a short sequence which will act as a template to prime extended DNA synthesis of telomeric DNA sequences on the leading strand. Further extension of the leading strand provides the necessary template for DNA polymerase (a) to complete synthesis of the lagging strand. This mechanism leaves the telomere itself with a protruding 3 end. In mammalian chromosomes, the single-stranded end is believed to loop round and invade the double helix several kilobases proximally, producing a triple-stranded structure resembling the mitochondrial D-loop, which is stabilized by binding telomere-specific proteins. However, the actual nature of the telomere sequence may not be important.

3. Helping establish the three-dimensional architecture of the nucleus and/or chromosome pairing. Chromosome ends appear to be tethered to the nuclear membrane, suggesting that telomeres help position chromosomes.

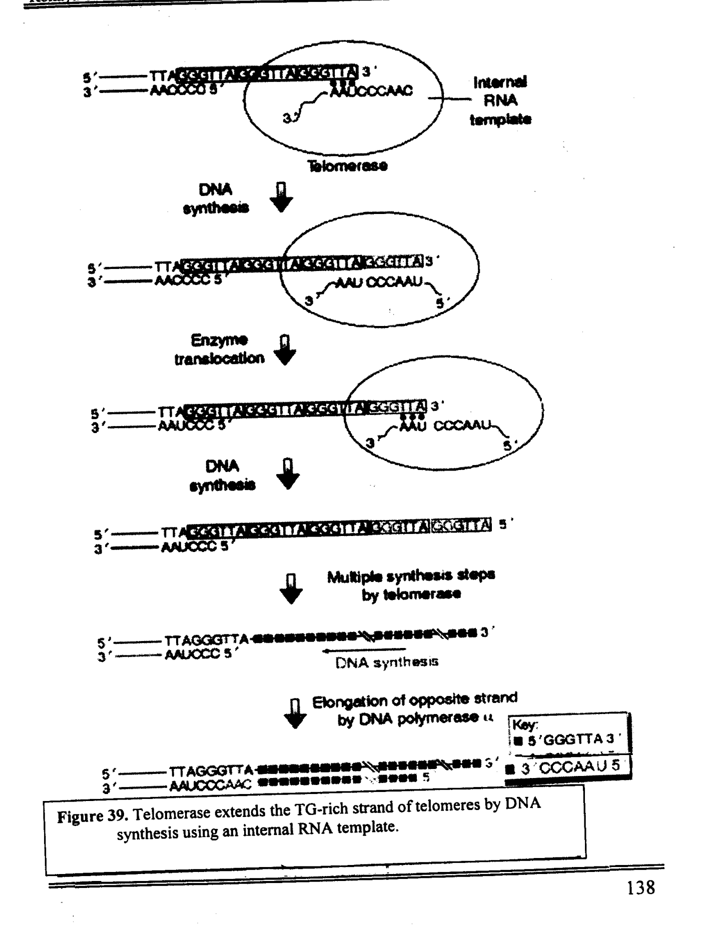


Figure 39. Telomerase extends the TG-rich strand of telomeres by DNA synthesis using an internal RNA template.

Just internal to the essential telomeric repeats, eukaryotic chromosomes also have a more complex set of repeats calledsubtelomeric or telomere-associated repeats. Their sequences are not conserved in eukaryotes and their function is unknown

Packaging of DNA into chromosomes

In the cell the structure of each chromosome is highly complicated. Even in the interphase nucleus the 2 nm DNA double helix is subject to at least two levels of coiling.

1. The most fundamental unit of packaging is the **nucleosome**. This consists of a central core of eight histone proteins, small highly conserved basic proteins of 102-135 amino acids. Each core comprises two molecules each of histones H2A, H2B, H3 and H4, around which a stretch of 146 bp of double-stranded DNA is coiled in 1.75 turns. Adjacent nucleosomes are connected by a short length of spacer DNA. Electron micrographs of suitable preparations show a 'string of beads' appearance.
2. The **string of beads**, approximately 10 nm in diameter, is in turn coiled into a **chromatin fiber** of 30 nm diameter. The interphase chromosome seems to consist of these chromatin fibers, probably organized into long loops.
3. During cell division, the chromosomes become ever more highly condensed.

4. **Central scaffold** The DNA in a metaphase chromosome is compacted to about 1/10, 000 of its stretched-out length. Loops of the 30 nm chromatin fiber, containing 20-100 kb of DNA per loop, are attached to a central scaffold. This consists of nonhistone acidic proteins, notably topoisomerase II, an enzyme which has the interesting ability to pass one DNA double helix through another by cutting a gap and repairing it. Topoisomerase II and some other chromatin proteins are known to bind to AT-rich sequences, and the chromatin loops may be attached by stretches of several hundred base pairs of highly AT-rich (>65%) DNA (scaffold attachment regions). In the chromatids of a metaphase chromosome the **loop-scaffold complex** is compacted yet further by coiling **(Fig.40)**.

Heterochromatin and euchromatin

In the interphase nucleus most of the chromatin (euchromatin) exists in an extended state, dispersed through the nucleus and staining diffusely. However, some chromatin remains highly condensed throughout the cell cycle and forms dark-staining regions (heterochromatin). Genes located in euchromatin may or may not be expressed, depending on the cell type and its metabolic requirements, but genes that are located within heterochromatin, either naturally or as the result of a chromosomal rearrangement, are very unlikely to be expressed. There are two classes of heterochromatin

- ♦ **Constitutive heterochromatin** is always inactive and condensed. It consists largely of repetitive DNA and is found in and around the centromeres of chromosomes and in certain other regions.

- ♦ **Facultative heterochromatin** can exist in either a genetically active (decondensed) or an inactive and condensed form, as in the case of mammalian X-chromosome inactivation.

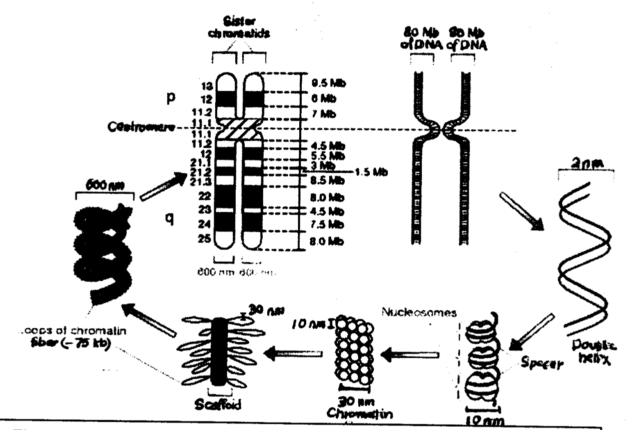


Figure 40. Diagram showing the different levels of DNA packaging. Notice the DNA starts out as single strand double helix and continues to be condensed until it reaches the chromosomal level.

Summary

Metaphase chromosomes as seen under the microscope in a dividing cell they are switched off DNA .
Interphase chromosome are switched on DNA , functioning interphase chromosome that controls cellular activities . it is extended and diffuse in the nucleus in the form of chromatin (Euchromatin and heterochromatin) . it comprises only a single chromatid and one DNA double helix, metaphase chromosome as an organelle function through three regions (1)functional centromere, (2) telomeres and (3) replication origin.

Centromere is single primary constriction, where the 2 sister chromatids are joined. Centromere forms pair of kinetochores linking the chromosome and the two spindle poles. It is essential for segregation of the 2 chromatids during cell division. Chromosome fragments that lack a centromere (acentric fragments) fail to be included either of the daughter cells. Structure and function of centromeres is related to specific DNA sequences which are very short and repetitive. They are of heterochromatin type. Centromere has four main functions: (1) Orientation of chromosomes at metaphase, (2) movement of chromosomes at anaphase, (3) formation of chromatids and (4) shape of chromosomes.

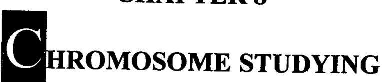
Origins of replication The DNA in replicates once per cell cycle. The initiation of replication lie at certain sites across the length of the DNA molecule (interphase chromosome). Replication can be initiated at multiple sites over regions which are tens of kilobases long to complete synthesis of DNA in short time.

Telomeres Telomeres are specialized structures, comprising DNA and protein, which cap the ends of chromosomes. It consists of a long array of tandem repeats. (TG-rich sequences). Its functions are (1) maintains the structural integrity of the chromosome, (2) ensure complete replication of the extreme ends of chromosomes and (3) establish the three-dimensional architecture of the nucleus.

The interphase chromosome (2 nm DNA double helix) is packed in the **nucleosome** (central core of eight histones proteins). Nucleosomes are coiled in 10 nm fiber which coiled into a chromatin fiber of 30 nm diameter.

Metaphase chromosome appears when the 30 nm chromatin attaches to central scaffold which forms loop-scaffold complex. Loop-scaffold complex coils more and more to form the metaphase chromosome.

CHAPTER 8



CHROMOSOME STUDYING

Introduction

Chromosomes can only be seen in the dividing cells, and obtaining dividing cells directly from the human body is difficult. Bone marrow is a possible source, but it is much easier all round to take an accessible source of non-dividing cells and culture them in the laboratory. Blood is the material of choice. Most people don't mind giving a few milliliters of blood, and the T lymphocytes in blood can be easily induced to divide by treatment with lectins such as phytohemagglutinin. Other common sources include fibroblasts grown from skin biopsies and (for prenatal diagnosis) chorionic villi or fetal cells shed into the amniotic fluid.

Mitotic chromosomes

Although chromosomes were described accurately in some organisms as early as the 1880s, for many decades all attempts to prepare spreads of human chromosomes produced a tangle that defied analysis. The key to getting analyzable spreads was a new technique, growing cells in liquid suspension and treating them with hypotonic saline to make them swell. This allowed the first good quality preparations to be made in 1956s. White cells from blood are put into a rich culture medium laced with phytohemagglutinin and allowed to grow for 48-72 hours, by

which time they should be dividing freely. Nevertheless, because mitotic (M) phase occupies only a small part of the cell cycle, few cells will be actually dividing at any one time. The mitotic index (proportion of cells in mitosis) is increased by treating the culture with a spindle disrupting agent such as colcemid or colchicine. Cells reach M phase of the cycle, but are unable to leave it, and so cells accumulate in metaphase of mitosis. Often it is preferable to study prometaphase chromosomes, which are less contracted and so show more detail.

Cell cultures can be prevented from cycling by thymidine starvation; when the block is released the cells progress through the cycle synchronously. By trial and error, the time after release can be determined when a good proportion of cells are in the desired prometaphase stage.

Meiotic chromosomes

Meiosis can only be studied in testicular or ovarian samples. Female meiosis is especially difficult, as it is active only in fetal ovaries, whereas male meiosis can be studied in a testicular biopsy from any post-pubertal male who is willing to give one. The results of meiosis can be studied by analyzing chromosomes from sperm, although the methodology for this is cumbersome. Meiotic analysis is used for some investigations of male infertility.

Identification of human Chromosomes

Until the 1970s chromosomes were identified on the basis of their size and the position of the centromeres. This allowed

chromosomes to be classified into groups but not unambiguously identified. **(Table 5).**

Table 5. Human chromosome groups:

Group	Chromosomes No.	Description
A	1- 3	Largest; 1 and 3 are metacentric but 2 is submetacentric
B	4,5	Large; submetacentric with two arms very different in size
C	6-12 +X	Medium size; submetacentric
D	13- 15	Medium size; acrocentric with satellites
E	15- 18	Small; 16 is metacentric but 17 and 18 are submetacentric
F	19,20	Small; metacentric
G	12,22,Y	Small; acrocentric, with satellites on 21 and 22 but not on the Y

The introduction of banding techniques finally allowed each chromosome to be identified, as well as permitting more accurate definition of translocation breakpoints, subchromosomal deletions, etc. Banding resolution can be increased by using more elongated chromosomes, for example chromosomes from prometaphase or earlier, rather than metaphase. Typical high-resolution banding procedures for human chromosomes can resolve a total of 400 to 550 or 850 bands.

The chromosome constitution (**karyotype**) is described by a that states the total number of chromosomes and the sex chromosome constitution. Human females and males are 46,XX and 46,XY respectively. When there is a chromosomal

abnormality the karyotype also describes the type of abnormality and the chromosome bands or sub-bands affected. Chromosomes are displayed as a karyogram (often loosely described as a karyotype). Karyograms prepared by cutting up a photograph of the spread, matching up homologous chromosomes and sticking them back down on a card - or nowadays more often by getting an image analysis computer to do the job **(Fig.41)**.

Chromosome banding

Various treatments involving denaturation and/or enzymatic digestion, followed by incorporation of a DNA-specific dye, can cause human and other mitotic chromosomes to stain as a series of light and dark bands. Banding patterns are interesting (as well as being useful to cytogeneticists) because they provide evidence of some sort of structure over 110 Mb regions. The banding patterns correlate with other properties. Regions that stain as dark G bands replicate late in S phase of the cell cycle and contain more condensed chromatin,

While R bands (light G bands) generally replicate early in S phase, and have less condensed chromatin. Genes are mostly concentrated in the R bands, while the later replicating, more condensed. G-band DNA is less active transcriptionally. There are also differences in the types of dispersed repeat elements found in G and R bands. The pattern of chromosome banding is highly specific in each chromosome of a species. First quinacrine dye was used for study of banding pattern. Subsequently, a number of dyes were identified which are able to produce chromosome bands.

Figure 41.

This is a karyogram of the best banding patterns that might be seen on each chromosome, and not a picture of how chromosomes appear in any one cell under the microscope.

Chromosomes are numbered in order of size, except that 21 is actually smaller than 22.

Arrays of repeated ribosomal DNA genes on the short arms of the acrocentric chromosomes 13, 14, 15, 21 and 22 often appear as thin stalks carrying knobs of chromatin (satellites).

Heterochromatin occurs at centromeres, on much of the Y chromosome long arm, at secondary constrictions on 1q, 9q and 16q, and on the short arms of the acrocentric chromosomes

There are four types of banding techniques, (1) Q bands, (2) G bands, (3) R bands and (4) C bands. These are briefly described below :

G Bands

In this technique, first the chromosomes are treated with saline. Detergents and urea may also be used for this purpose. The chromosomes are then treated with Giemsa stain **(Fig. 42).**The slides of such chromosomes are examined for banding pattern under light microscope. G bands appear to reflect variation of protein sulphur in different regions of a chromosome. G Bands similar to bands are AT-rich DNA.

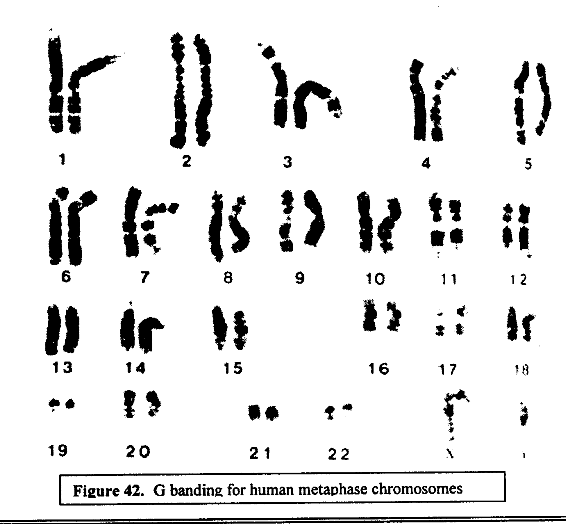


Figure 42. G banding for human metaphase chromosomes

Q Bands

In this technique, the chromosomes are treated with quinacrine mustard, quinacrine or Hoechst 33258 or some other dyes. Then the chromosomes are examined for banding pattern under fluorescence microscope. Some regions show intensely fluorescent bands. These regions are considered to be rich in AT. The less fluorescent bands indicate regions which are rich in GC.

R Bands

In this technique, first the chromosomes are treated in a buffer at a high temperature. These chromosomes are then treated with Giemsa stain. This gives altogether a different pattern known as R banding. The bands are reverse of the G bands.

C Bands

The chromosomes are first treated with a moderately strong alkali and then by warm saline solution. The chromosomes are then stained with Giemsa dye. In this technique, the region around the centromere which contains satellite DNA exhibits dark stain under microscope.

The banding pattern for a given chromosome is constant but different with different banding procedures. The banding pattern studies are useful in the identification of chromosomes in three different ways :

1. Such studies help in identification of individual chromosome of a species with certainty because banding pattern is highly specific for a chromosome.
2. It helps in identification of structural chromosomal changes, deletion, duplication, translocation and inversion.
3. It is also useful in assigning various linkage groups to specific chromosome and in accurate gene mapping.

Role of chromosomes in heredity

Chromosomes are considered as physical basis of inheritance. The first conclusive evidence that chromosomes carry the units of inheritance was put forward by Sutton in 1903. working with grasshopper, he gave a hypothesis that chromosomes contain genes and their behavior during meiosis is the physical basis of Mendelian laws of heredity. Thus his work formed the basis of chromosomal theory of heredity. Now this theory is universally accepted. Various evidences which support that genes are located on the chromosomes which from the physical basis of heredity are briefly presented below :

1. Similarities between chromosomes and genes. There are some similarities between chromosomes and genes, (1) two copies of each in somatic cells on copy in gametic cell, (2) self duplication or replication capacity, (3) segregation during meiosis and (4) mutability. All these parallel features between gene and chromosomes suggest that chromosomes carry genes and represent the physical basis of heredity.
2. Studies on Sex Chromosomes. Chromosomes play an important role in sex determination. In unisexual diploid

organisms, there are two sex chromosomes and rest are autosomes. In case of Drosophila and man the sex chromosomes are of X and Y type. The XX sex is female and XY is male. This has also proved beyond doubt that genes are located on the chromosomes.

3. Linkage Studies. The linkage studies of Morgan in Drosophila clearly demonstrated that genes are located on the chromosomes in a linear fashion and genes of one chromosome are linked together, if crossing over does not occur.
4. Studies on structural chromosomal changes. Studies on structural chromosomal changes especially deficiency and duplication suggest that genes are located in chromosomes. Because loss of some part of a chromosome leads to alteration in a specific character. Similarly, duplication of chromosome segment affects the phenotypic expression of a particular trait. This provides strong evidence in favors of chromosomal theory of inheritance. Translocations and inversions lead to change in the normal sequence of genes. Such changes in gene sequence are reflected through the pairing of translocated or inverted chromosomes during meiosis which is observable cytologically. This provides strong support in favor of chromosomal theory of inheritance.
5. Monosomic and nullisomic analysis. Studies of monosomic and nullisomic analysis are useful in locating genes on different chromosomes. Loss of one chromosome from one set leads to alteration of some characters in an individual indicating the presence of genes on the missing chromosome. Nullisomics, if viable are also useful in

assigning the genes to different chromosomes of an individual. Monosomic series can be developed for all the chromosomes of an individual and genes for different characters can be easily located.

6. Studies on crossing over .Cytological studies on Drosophila and proved that crossing over occurs due to exchange of segments between homologous chromosomes. This clearly demonstrated that genes are located in chromosomes .The concept that crossing over depends on the distance between two gene further supports the chromosomal of inheritance.

7. Studies on Bar locus. The studies on bar locus of drosophila provides strong evidence that genes are present in the chromosome. Because single dose of bar segment produces normal eye, double dose leads to bar eye and triple dose results in ultra bar eye.

8. Studies on polyploids. The polyploidy individuals have more than two copies of each chromosome and thus exhibit complex inheritance. For example, allotetraploids have four copies of each type of chromosome. Cytological studies of such individuals show quadrivalent formation during meiosis which indicates that each gene also has four copies in such individuals.

9. Biochemical studies. Biochemical studies reveal that hereditary units (genes) are composed of DNA in eukaryotes and RNA in some prokaryotes. The major part of DNA is found in chromosomes which proves beyond doubt that chromosomes are the carriers of hereditary units what we call genes.

10. Studies on non disjunction of chromosomes also reveal that genes are located on the chromosomes.
11. Transformation and transduction studies have clearly brought out that genes are located on the chromosomes and are composed of DNA in eukaryotes and RNA in some prokaryotes.

The above evidences clearly show that genes are located on the chromosomes and thus chromosomes are the physical carriers of hereditary units which are responsible for transmission of characters from one generation to other generation.

Summary

To study the metaphase chromosomes must be seen in the dividing cells, and obtaining dividing cells directly from the human body is difficult. Sources of dividing cells are: bone marrow, blood, fibroblasts from skin and chorionic villi. while the meiotic chromosomes studied in testicular or ovarian samples

Identification of human chromosomes

Before 1970s chromosomes were identified on the basis of their size position of centromeres and they classified into 8 groups named from A to G (group A, B→ G).

After 1970, the introduction of banding allowed each chromosome to be identified. Banding resolution (number of bands per each chromosome) can be increased by using more elongated chromosomes, Typical high-resolution banding procedures for human chromosomes can resolve a total of 400 to 550 or 850 bands per karyotype.

Karyotype: The chromosome constitution) is described by that states the total number of chromosomes and the sex chromosome. Human females and males are 46,XX and 46,XY respectively.

Chromosome banding

After denaturation or enzymatic digestion, of the metaphase chromosomes they stained as a series of light and dark bands with DNA-specific dye, mitotic

G Bands: The chromosomes are treated with saline. Digestion by trypsine or urea. The chromosomes are then stained with Giemsa stain G bands appear to reflect variation of protein sulphur in different regions of a chromosome.

Q Bands: In this technique, the chromosomes are treated with quinacrine mustard, quinacrine or Hoechst 33258 and examined for under fluorescence microscope.

R Bands: The chromosomes are treated high temperature then stained with Giemsa stain.

C Bands: The chromosomes are treated with a moderately strong alkali and then by warm saline solution. The chromosomes are then stained with Giemsa dye. By this technique we can examine the dark stained centromeric region which contains satellite DNA.

Role of chromosomes in heredity

Chromosomes are considered as physical basis of inheritance. There are more than 10 evidences reverse this suppose to a fact. The evidences clearly show that genes are located on the chromosomes and thus chromosomes are the physical carriers of hereditary units which are responsible for transmission of characters from one generation to other generation.

CHAPTER 9

CHROMOSOME ABNORMALITIES

Introduction

Chromosome abnormalities might be defined as changes resulting in a visible alteration of the chromosomes. How much can be seen depends on the technique used. The smallest loss or gain of material visible by traditional methods on standard cytogenetic preparations is about 4 megabases of DNA. However, fluorescence in situ hybridization allows much smaller changes to be seen; the development of molecular cytogenetics has removed any clear dividing line between changes described as chromosomal abnormalities and changes thought of as molecular or DNA defects.

An alternative definition of a chromosomal abnormality is an abnormality produced by specifically chromosomal mechanisms. Most chromosomal aberrations are produced by misrepair of broken chromosomes, by improper recombination or by malsegregation of chromosomes during mitosis or meiosis.

Types of chromosomal abnormality:

Constitutional: A chromosomal abnormality must have been present very early in development, most likely the result of an abnormal sperm or egg, or may be abnormal fertilization or an

abnormal event in the early embryo. It is present in all cells of the body.

Somatic or acquired: present in only certain cells or tissues. An individual with a somatic abnormality is a mosaic , containing cells with two different chromosome constitutions, with both cell types deriving from the same zygote.

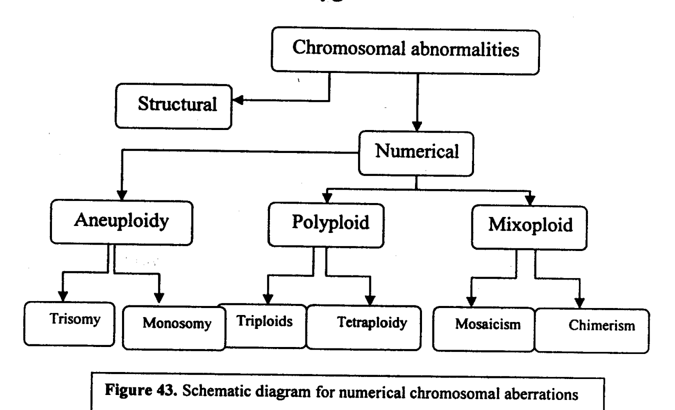


Figure 43. Schematic diagram for numerical chromosomal aberrations

Chromosomal abnormalities, whether constitutional or somatic, mostly fall into two categories: **Numerical and structural** abnormalities.

Numerical abnormalities:

Numerical chromosomal abnormalities involve gain or loss of complete chromosomes. Three classes of numerical

chromosomal abnormalities can be distinguished: polyploidy, aneuploidy and mixoploidy **(Fig, 43)**.

1- Polyploidy:

Triploids (Table 6). The most usual cause of triploids is fertilization of a single egg by two sperms (dispermy). Sometimes the cause is a diploid ovum fertilized by one sperm or fertilization of normal ovum by diploid sperm **(Fig.44, A,B and C)**. It is very seldom survive to term, and the condition is not compatible with life.

Table 6. Nomenclature of chromosome abnormalities

Numerical abnormalities	
Triploidy	69,XXX, 69,XXY, 69,XYY
Tetraploidy	92,XXXX, 92,XXYY
Trisomy	e.g. 47,XX, +21
Monosomy	e.g. 45,X
Mosaicism	e.g. 47,XXX / 46,XX
Structural abnormalities:	
Deletion	e.g. 46,XY, del(4)
Inversion	e.g. 46,XY, del(4) (p16.3) 46,XX, del(5) (q13q33)
Duplication	e.g. 46,XX, dup(1) (q22q25)
Insertion	e.g. 46,XX, ins(2) (p13q21q31)
Ring	e.g. 46,XY, r(7) (p22q36)
Marker	e.g. 47,XX, +mar
Translocation, reciprocal	e.g. 46,XX, t(2;6) (q35;p21.3)

Tetraploidy is much rarer and always lethal. It is usually due to failure to complete the first zygotic division: the DNA has replicated to give a content of 4C, but cell division has not then taken place as normal **(Fig. 44, D)**. Although constitutional

polyploidy is rare and lethal, all normal people have some polyploid cells.

2- Aneuploidy

Euploidy means having complete chromosome sets (n, 2n, 3n, etc.). **Aneuploidy** is the opposite, that is, one or more individual chromosomes extra or missing from the euploid set

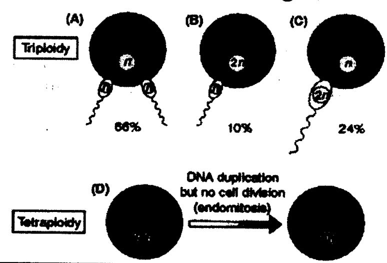


Figure 44. Causes of triploidy
(A) by fertilization of a single egg by two sperm
(B) diploid egg fertilized by normal sperm
(C) normal egg fertilized by diploid sperm.
(D) Tetraploidy results from failure of the first mitotic division after fertilization.

Trisomy means having three copies of a particular chromosome in a diploid cell, for example trisomy 21 (47,XX or XY, +21) in Down syndrome.

Monosomy is the corresponding lack of a chromosome, for example monosomy X (45,X) in Turner syndrome. Cancer cells often show extreme aneuploidy, with multiple chromosomal abnormalities.

Autosomal monosomies have even more catastrophic

consequences than trisomy. These abnormalities must be the consequence of an imbalance in the levels of gene products encoded on different chromosomes. Normal development and function depend on innumerable interactions between gene products, including many that are encoded on different chromosomes. Altering the relative numbers of chromosomes will affect these interactions.

Mechanisms of aneuploidy:

- **Nondisjunction:**

In 1st miotic division, failure of paired chromosomes to separate (disjoin).

In 2nd meiotic division failure of sister chromatids to disjoin.

Nondisjunction in meiosis produces gametes with 22 or 24 chromosomes, which after fertilization by a normal gamete make a trisomic or monosomic zygote.

In mitiotic divition: the cause is failure of the two-sister chromatids to separate. Nondisjunction in mitosis produces a mosaic tissue cells.

- **Anaphase lag**: failure of a chromosome or chromatid to be incorporated into one of the daughter nuclei following cell division, as a result of delayed movement (lagging) during anaphase. Chromosomes that do not enter a daughter cell nucleus are lost.

3- Mixoploidy

Mosaicism an individual possesses two or more genetically different cell lines all derived from a single zygot. It is compatible with life. Aneuploidy mosaics are common. For example, mosaicism resulting in a proportion of normal cells and a proportion of aneuploid (e.g. trisomic) cells can be ascribed to nondisjunction or chromosome lag occurring in one of the mitotic divisions of the early embryo (any monosomic cells that are formed usually die out). **(Fig.45)**Polyploidy mosaics (e.g. human diploid/triploid mosaics) are occasionally found. As gain or loss of a haploid set of chromosomes by mitotic nondisjunction is most unlikely, human diploid/triploid mosaics most probably arise by fusion of the second polar body with one of the cleavage nuclei of a normal diploid zygote.

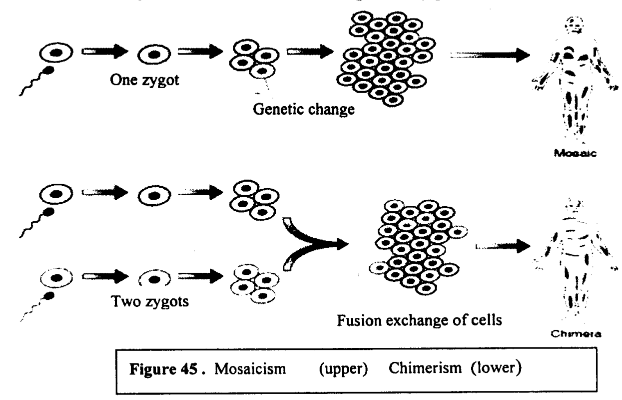


Figure 45. Mosaicism (upper) Chimerism (lower)

Chimerism an individual has two or more genetically different cell lines originating from different zygotes **(Fig.45)**. Abnormalities that are would be lethal.

Clinical consequences of numerical abnormalities:

Having the wrong number of chromosomes has serious, usually lethal, consequences.

Having the wrong number of sex chromosomes has fewer ill effects than having the wrong number of any autosome. 47,XXX and 47,XYY people often function within the normal range; 47,XXY men have relatively minor problems compared to people with any autosomal trisomy, and even monosomy, in 45,X women, has remarkably few major consequences.

In fact, since normal people can have either one or two X chromosomes, and either no or one Y, there must be special mechanisms that allow normal function with variable numbers of sex chromosomes. In the case of the Y chromosome, this is because it carries very few genes, whose only important function is to determine male sex. For the X chromosome, the special mechanism of lyonization. controls the level of X-encoded gene products independently of the number of X chromosomes present in the cell.

Autosomal monosomy is invariably lethal at the earliest stage of embryonic life. On every chromosome there are probably a few genes where a halving of the level of the gene product is incompatible with development. Also, while such a halving is not obviously pathogenic for most genes, it may have minor effects, and the combination of hundreds or thousands of

these minor effects could be enough to disrupt normal development of the embryo. Trisomies make a smaller change than monosomies in relative levels of gene products, and their effects are somewhat less. Trisomic embryos survive longer than monosomic ones, and trisomies 13, 18 and 21 are compatible with survival until birth. Interestingly, these three chromosomes seem to be relatively poor in genes.

It is not so obvious why triploidy is lethal in humans and other animals. With three copies of every autosome, the dosage of autosomal genes is balanced and should not cause problems. Triploids are always sterile because triplets of chromosomes cannot pair and segregate correctly in meiosis, but many triploid plants are in all other respects healthy and vigorous. The lethality in animals is probably explained by imbalance between products encoded on the X chromosome and autosomes, which lyonization is unable to compensate.

Causes of structural chromosomal abnormalities:

Chromosome breaks may occur either as a result of:

1. Damage to DNA (e.g. by radiation or chemicals)
2. As part of the mechanism of recombination.

In G_2 phase of the cell cycle chromosomes consist of two chromatids. Breaks occurring at this stage are manifest as chromatid breaks, affecting only one of the two sister chromatids. Breaks occurring in G_1 phase, if not repaired before S phase, appear later as chromosome breaks, affecting both chromatids. Cells have enzyme systems that recognize and if possible repair broken chromosome ends. Repair can be either by joining two

broken ends together, or by capping a broken end with a telomere. Cell cycle checkpoint mechanisms normally prevent cells with unrepaired chromosome breaks from entering mitosis; if the damage cannot be repaired, the cell commits suicide (apoptosis).

Any resulting chromosome that has no centromere (acentric) or two centromeres (dicentric) will not segregate stably in mitosis, and will eventually be lost. Chromosomes with a single centromere can be stably propagated through successive rounds of mitosis, even if they are structurally abnormal.

Meiotic recombination between mispaired chromosomes **(Table.7)** is a common cause of translocations, especially in spermatogenesis.

Isochromosomes An additional rare class of structural abnormality is isochromosomes. These are symmetrical chromosomes consisting of either two long arms or two short arms of a particular chromosome. They are believed to arise from an abnormal U-type exchange between sister chromatids just next to the centromere of a chromosome. Isochromosomes are rare except for i(Xq); i(21q) are an occasional cause of Down syndrome.

Balanced structural chromosomal abnormalities are balanced if there is no net gain or loss of chromosomal material.

Unbalanced if there is net gain or loss. Unbalanced abnormalities can arise directly, through **deletion** or, rarely, **duplication**, or indirectly by malsegregation of chromosomes during meiosis in a carrier of a balanced abnormality.

In general, balanced abnormalities (inversions, balanced translocations) have no effect on the phenotype, although there are important **exceptions** to this:

Table 7. Misrepair of breaks or recombination between nonhomologous chromosomes

No. of breaks	One chromosome involved	Two chromosomes involved
One break	Terminal deletion (healed by adding telomere)	
Two breaks	Interstitial deletion;	Reciprocal translocation
	Inversion;	Robertsonian translocation
	Ring chromosome	Duplication or deletion by unequal recombination
	Duplication or deletion by unequal sister-chromatid exchange	
Three breaks	Various rearrangements, e.g. inversion with deletion, intrachromosomal insertion	Interchromosomal insertion (direct or inverted)

- A chromosome break may disrupt an important gene (**Fig. 46).** The break may affect expression of a gene even though it does not disrupt the coding sequence. It may separate a gene from a control element, or it may put the gene in an inappropriate chromatin environment, for example translocating a normally active gene into heterochromatin.

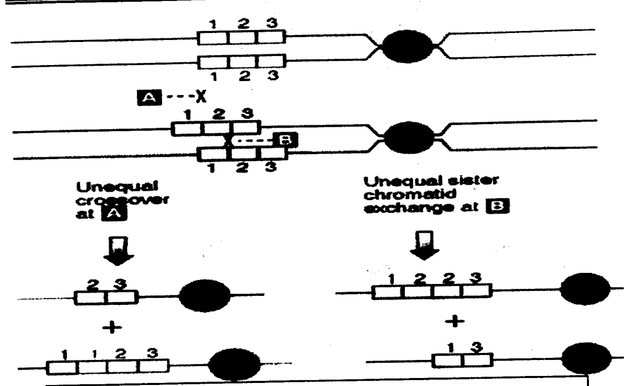


Figure 46.
A, Unequal crossover involves unequal pairing of nonsister chromatids followed by chromatid breakage and rejoining.
B, Unequal sister chromatid exchange involves unequal pairing of sister chromatids followed by chromatid breakage and rejoining.

- Balanced X-autosome translocations cause problems with X-phenotype.
- **Robertsonian translocations** are sometimes called **centric fusions**, but this is misleading because in fact the **breaks** are in the proximal short arms. The translocation chromosome is really **dicentric**, but because the two centromeres are very close together they function as one, and the chromosome segregates regularly. The distal parts of the two short arms are lost as an **acentric fragment**. Short arms of acrocentric chromosomes contain only arrays of repeated ribosomal RNA genes, and the loss of

two short arms has no phenotypic effect. Because there is no phenotypic effect, Robertsonian translocations are regarded as balanced, even though in fact some material has been lost.

Carriers of balanced structural abnormalities can run into trouble during meiosis:

- Carrier of a balanced **reciprocal translocation** can produce gametes that after fertilization give rise to an entirely normal child, a phenotypically normal balanced carrier, or various unbalanced karyotypes that always combine monosomy for part of one of the chromosomes with trisomy for part of the other **(Fig.46)**.
- A carrier of a balanced **Robertsonian translocation** can produce gametes that after fertilization give rise to an entirely normal child, a phenotypically normal balanced carrier, or a conceptus with full trisomy or full monosomy for one of the chromosomes involved
- produce unbalanced offspring because when the inverted and non-inverted homologs pair they form a loop so that matching segments pair along the whole length of the chromosomes. If a crossover
- Carrier of **Paracentric inversions** form similar loops, but any crossover within the loop generates an acentric or dicentric chromosome, which is unlikely to survive **(Fig.47)**.

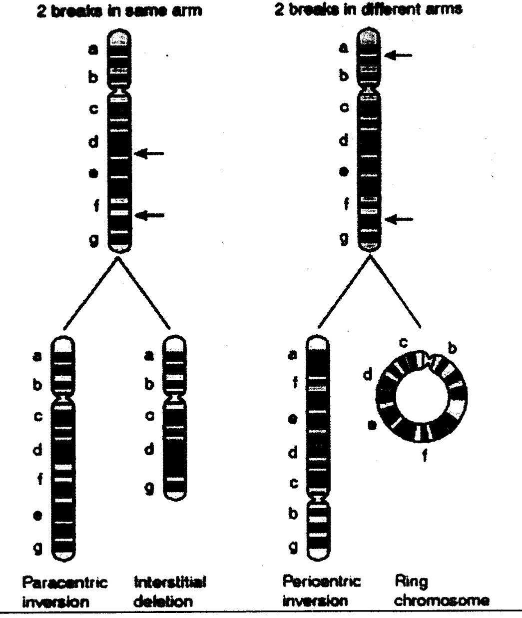


Figure 47. Possible stable results of two breaks on a single chromosome.

- A carrier of a **pericentric inversion (Fig.47)** may occurs within the loop, the result is a chromosome carrying an unbalanced deletion and duplication.

The cause is thought to be trisomy rescue: a conceptus that is trisomic and would otherwise die, occasionally loses one chromosome by mitotic nondisjunction or anaphase lag from a totipotent cell. The euploid progeny of this cell form the embryo, while all the aneuploid cells die. If each of the three copies has an equal chance of being lost, there will be a two in three chance of a single chromosome loss leading to the normal chromosome constitution and a one in three chance of uniparental disomy (either paternal or maternal). Uniparental isodisomy may possibly arise by selection pressure on a monosomic embryo to achieve euploidy by selective duplication of the monosomic chromosome.

Summary

Chromosomal abnormality produced by specifically chromosomal mechanisms. Most chromosomal aberrations are produced by misrepair of broken chromosomes, by improper recombination or by malsegregation of chromosomes during mitosis or meiosis.

Constitutional abnormality: A chromosomal abnormality present early in development, result of an abnormal sperm or egg, or abnormal fertilization or an abnormal event in the early embryo. It is present in all cells of the body.

Somatic or acquired abnormality: present in certain cells or tissues. An individual with a somatic abnormality is a mosaic, containing cells with two different chromosome constitutions, with both cell types deriving from the same zygote.

Structural abnormalities is gain or loss of certain part of particular chromosome (s).

Numerical: chromosomal abnormalities involve gain or loss of complete chromosomes.

1- Polyploidy:

Triploids caused by (1) fertilization of a single egg by two sperms. (2) diploid ovum fertilized by one sperm.

Tetraploidy. It is usually due to failure to complete the first zygotic division (lethal).

2- Aneuploidy is one or more individual chromosomes extra or missing from the euploid set.

Trisomy means having three copies of a particular chromosome e.g. trisomy 21 in Down syndrome.

Monosomy is lack of a chromosome e.g. monosomy X (45,X) in Turner syndrome.

Mechanisms of aneuploidy:

- **Nondisjunction:** In 1st miotic division, failure of paired chromosomes to separate (disjoin). In 2nd Meiotic divition failure of sister chromatids to disjoin.

- **Anaphase lag**: failure of a chromosome or chromatid to be incorporated into one of the daughter nuclei following cell division,

3- Mixoploidy

Mosaicism an individual possesses two or more genetically different cell lines all derived from a single zygot..

Chimerism an individual has two or more genetically different cell lines originating from different zygotes (lethal).

Clinical consequences of numerical abnormalities:

Having the wrong number of autosomal chromosomes has serious, usually lethal, consequences. Having the wrong number of sex chromosomes has fewer ill effects.

Causes of structural chromosomal abnormalities:

Chromosome breaks may occur either as a result of

1. Damage to DNA (by radiation or chemicals)
2. As part of the mechanism of recombination.

Balanced structural chromosomal abnormalities are balanced if there is no net gain or loss of chromosomal material, balanced abnormalities (inversions, balanced translocations) have no effect on the phenotype, with some **exceptions**:

- A chromosome break may disrupt an important gene

- Balanced X-autosome translocations cause problems with X-phenotype.

- Robertsonian translocations are sometimes called **centric fusions**

Unbalanced if there is net gain or loss of chromosomal material which affect on the phenotype.

CHAPTER 10

RECOMBINANT TECHNOLOGY

Introduction

Before DNA cloning, our knowledge of DNA was extremely limited. DNA cloning technology changed all that and revolutionized the study of genetics. Why was DNA cloning such an important technological advance? One important consideration is the tremendous size and complexity of DNA sequences (compared to protein sequences). Individual nuclear DNA molecules contain hundreds of millions of nucleotides. When DNA is isolated from cells using standard methods, these huge molecules are fragmented by shear forces, generating complex mixtures of still very large DNA fragments (typically 50-100 kb long). **(Fig. 48).**

The fundamentals of current DNA technology are very largely based on two quite different approaches to studying specific DNA sequences within a complex DNA population:

1. **DNA cloning:** The desired fragment must be selectively amplified so that it is purified essentially to homogeneity. Thereafter, its structure and function can be comprehensively studied, for example by DNA sequencing, in vitro expression studies, etc., and various manipulations can be achieved to change its structure by in vitro mutagenesis.

2. **Molecular hybridization**: The fragment of interest is not amplified, but instead is specifically detected within a

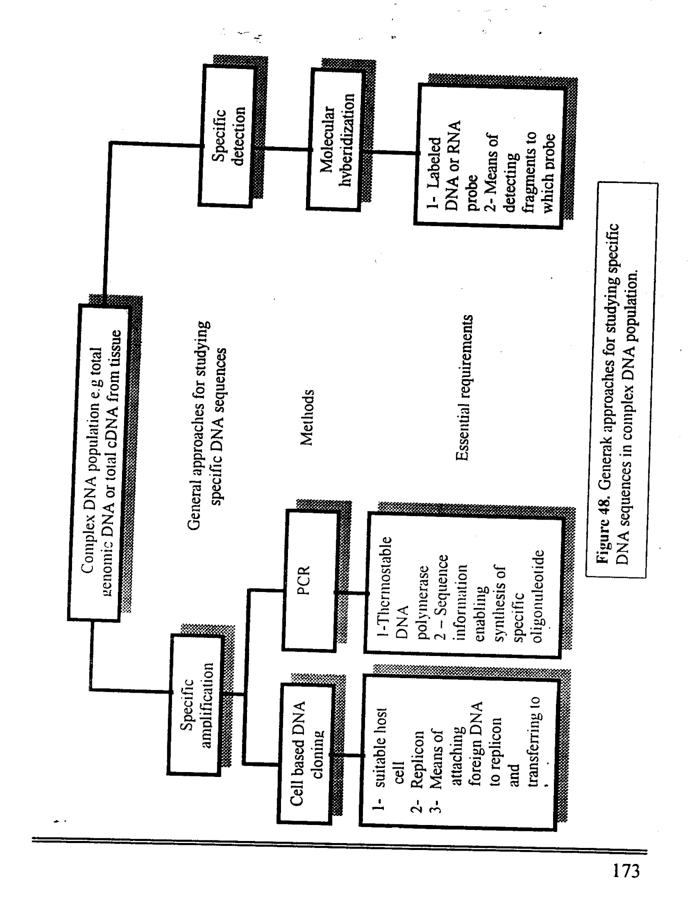


Figure 48. Generak approaches for studying specific DNA sequences in complex DNA population.

complex mixture of many different sequences. Its chromosomal location can be determined in this way and some information can be gained regarding its structure. If expressed, the sequence of interest can be detected within a complex RNA or cDNA population from specific cells, enabling comprehensive analysis of its expression patterns.

Given the above, how can relatively homogeneous DNA populations be prepared from such a complex starting mixture. In the case of DNA from wide variety of complex eukaryotic cells, one early approach had been to separate different classes of DNA fragments according to their base composition.

Ultracentrifugation The DNA preparation was submitted to ultracentrifugation in equilibrium density gradients (e.g. in CsCl density gradients). When this was achieved, the DNA was fractionated into a major band (the bulk DNA) and several minor **satellite DNA** bands. The satellite DNA species have different buoyant densities to the bulk DNA because they have unusual sequences and their base composition is different to the majority of the DNA in cells. These properties in turn reflect the involvement of satellite DNA in specific aspects of chromosome structure and function. Although valuable and interesting, the purified satellite DNAs were, however, a minor component of the genome and did not contain genes.

Complementary DNA (cDNA)

In order to study genes, methods had to be developed to purify them. Because mammalian genomes are complex, any specific gene or DNA fragment of interest normally represents only a tiny fraction of the total DNA in a cell. For example, the b-globin

gene comprises only 0.00005% of the 3300 megabases (Mb) of human genomic DNA, and even the massive 2.5 Mb dystrophin gene, the largest gene that has been identified, accounts for only about 0.08% of human genomic DNA.

One way of enriching for gene sequences is to isolate total RNA, or poly(A)$^+$ messenger RNA (mRNA) from suitable cells and convert this to complementary DNA (cDNA) using the enzyme reverse transcriptase. In some cases, this can result in a profound enrichment for specific exonic DNA sequences when the relevant genes are known to be expressed at very high levels in a specific cell type. In most cases, however, the desired gene sequences still represent only a tiny proportion of the total cDNA population.

DNA cloning

DNA cloning is a general method of selectively amplifying DNA sequences to generate homogenous DNA populations. In order to have a general method of studying a specific DNA sequence within a complex DNA population, the technology of DNA cloning was developed. The essential characteristic of DNA cloning is that the desired DNA fragments must be selectively amplified in some way, resulting in a programmed large increase in copy number of selected DNA sequences. In practice, this involves multiple rounds of DNA replication catalyzed by a DNA polymerase acting on one or more types of template DNA molecule. Essentially two different DNA cloning approaches are used:

- **Cell-free DNA cloning (PCR)**: The polymerase chain reaction (PCR) is a newer form of DNA

cloning which is enzyme mediated and is conducted entirely in vitro. (discussed latter).

- **Cell-based DNA cloning (recombinant DNA)**. This was the first form of DNA cloning to be developed, and is an in vivo cloning method. Cell-based DNA cloning has been used widely as a tool for producing quantities of pure DNA for physical characterization and functional studies of individual genes, gene clusters or other DNA sequences of interest, The first step in this approach involves attaching foreign DNA fragments in vitro to DNA sequences which are capable of independent replication. The recombinant DNA fragments are then transferred into suitable host cells where they can be propagated selectively. In the past, the term DNA cloning has been used exclusively to signify this particular approach **(Fig. 49).**

Target DNA

Any double strand DNA of suitable long and of unknown sequence can to be the target of cell-based cloning.

Vectors (Replicons)

Cell cloning requires that the foreign DNA fragments which are introduced into a host cell must be able to replicate. If not, the foreign DNA would soon be diluted out as the host cell undergoes many rounds of cell division. However, foreign DNA fragments will generally lack an origin of replication that will function in the host cell. They require, therefore, to be attached to

an independent replicon so that their replication is controlled by the replicon's origin of replication.

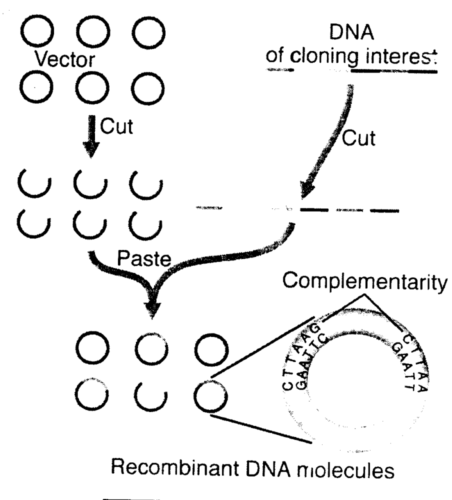


Figure. 49. DNA cloning in bacterial cells.

Replicons:

Could be provided in two ways.

1. **Cell chromosome:** One possibility is to introduce the fragments into the host cells chromosome and are propagated under the control of a host cell replicon. This is the way that retroviruses integrate their DNA into host cell

chromosomes. This approach suffers from numerous disadvantages, including the difficulty in retrieving the inserted DNA.

2. **Extra-chromosomal replicons:** The foreign DNA fragments are attached in vitro to a purified replicon and the resulting hybrid molecules are transferred into host cells, where they replicate independently of the host cell chromosomes. Because the foreign DNA fragments (target DNA) can be viewed as passengers of the replicon, replicons used for cloning are described as vector molecules. Notice that, although the resulting replication of the introduced DNA fragments is independent of the host chromosomes, the vector may have an origin of replication that originates from either a natural extra-chromosomal replicon or, in some cases, a chromosomal replicon (as in the case of yeast artificial chromosomes).

Extra-chromosomal replicons (vectors)

Basic types of extra-chromosomal replicons are:

1. **Plasmids** are small circular double-stranded DNA molecules which individually contain very few genes. Their existence is intracellular, being vertically distributed to daughter cells following host cell division, but they can be transferred horizontally to neighboring cells during bacterial conjugation. Natural examples include plasmids which carry the sex factor (F) and those which carry drug-resistance genes.

2. **Bacteriophages** are viruses which infect bacterial cells. DNA-containing. Bacteriophages often have genomes

containing double-stranded DNA which may be circular or linear. Unlike plasmids, they can exist extracellularly. The mature virus particle (**virion**) has its genome encased in a protein coat so as to facilitate adsorption and entry into a new host cell. After the virion attaches to the bacterial cell, the coat protein is discarded and the DNA is injected into the cell. **(Fig. 50)**.



Figure 50. Electron micrograph for bacteriophage infecting bacterial cell

3. **Cosmid vectors**. Cosmid vectors contain cos sequences inserted into a small plasmid vector. Large (~3044 kb) foreign DNA fragments can be cloned using such vectors in an in vitro packaging reaction because the total size of the cosmid vector is usually only about 8 kb.

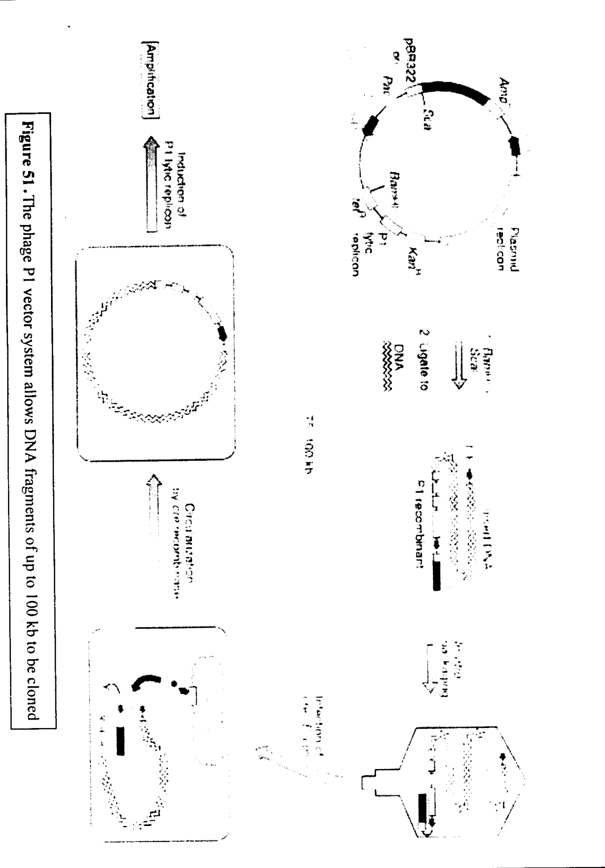


Figure 51 . The phage P1 vector system allows DNA fragments of up to 100 kb to be cloned

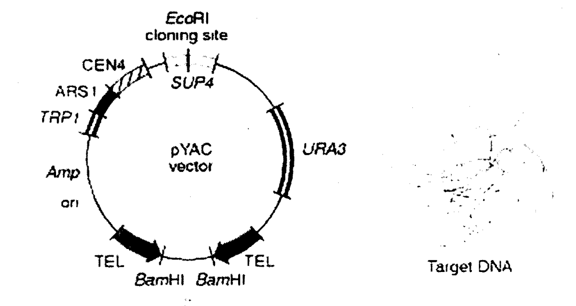
4. **BACs (Bacterial artificial chromosome) vectors:** Many vectors which are popularly used for DNA cloning in bacterial cells contain high to medium copy number replicons. However, an important disadvantage is that such vectors often show structural instability of inserts. However, because the resulting bacterial artificial chromosomes (BACs) contain a low copy number replicon, only very low yields of recombinant DNA can be recovered from the host cells.

5. **PACs(P1-derived artificial chromosome):**An improvement on the size range of inserts accepted by the basic P1 cloning system has been the use of bacteriophage T4 in vitro packaging systems with P1 vectors which enables the recovery of inserts up to 122 kb in size. More recently, features of the P1 and F-factor systems have been combined to produce P1-derived artificial chromosome (PAC) cloning systems **(Fig. 51).**

6. **YACs (yeast artificial chromosomes):**The most popularly used system for cloning very large DNA fragments involves yeast artificial chromosomes (**YACs**). Cloning in yeast cells offers, some advantages over cloning in bacterial cells. Certain eukaryotic repeated sequence are impossible, to propagate in bacterial cells but would be anticipated to be tolerated in yeast cells. However, the main advantage offered by YACs has been the ability to clone very large DNA fragments **(Fig. 52)**.

7. **Phagemid,** are plasmids which have been artificially manipulated so as to contain a small segment of the genome of a filamentous phage, such as M13, The selected

phage sequences contain all the cis-acting elements required for DNA replication and assembly into phage particles. They permit successful cloning of inserts several kilobases long.



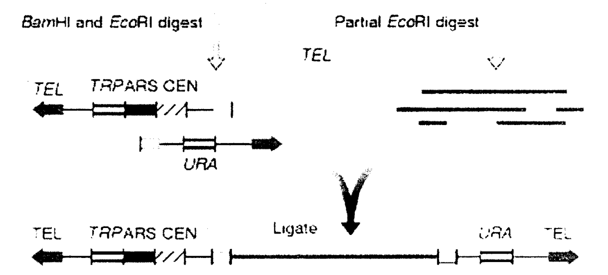


Figure 52. Making YACs. Vector DNA sequences include: CEN1, centromere sequence; TEL, telomere sequences; ARS1, autonomous replicating sequence; *Amp*, gene conferring ampicillin-resistance; ori, origin of replication for propagation in an *E. coli* host.

Typically, cloning systems are designed to ensure that joining of the foreign DNA fragment occurs at a unique location in the vector molecule. Additionally, they have in-built selection systems so that cells which contain the relevant vector molecule can be specifically selected. In many cases, there are additional screening systems to ensure detection and propagation of cells containing recombinant DNA.

Host cells

Although some DNA cloning systems involve human and other mammalian cells as hosts, the great bulk of cell-based DNA cloning has used modified bacterial or fungal host cells. Bacterial cells are widely used because of their capacity for rapid cell division. Bacterial cell hosts have a single circular double-stranded chromosome with a single origin of replication. Replication of the host chromosome subsequently triggers cell division so that each of the two resulting daughter cells contains a single chromosome like their parent cell.

However, the replication of extrachromosomal replicons is not constrained in this way: many such replicons go through several cycles of replication during the cell cycle and can reach high copy numbers. The host cells that are used for cloning are specialized cells whose genotype has been selected to optimize their use in DNA cloning.

Restriction endonucleases

A major boost to the development of cell-based DNA cloning was the discovery and exploitation of type II restriction

endonucleases, enzymes which normally cleave DNA whenever a small, specific recognition sequence, usually 48 base pairs (bp) long occurs.

The recognition sequences for the vast majority of type II restriction endonucleases are normally palindromes, that is the sequence of bases is the same on both strands when read in the 5'→3' direction.

Blunt ends: In some cases, the cleavage points occur exactly on the axis of symmetry, giving products (restriction fragments) which are blunt-ended.

Sticky ends: Overhanging ends generated by cleavage with a restriction nuclease are often described as sticky ends or cohesive termini because the two overhanging ends of each fragment are complementary in base sequence, and will have a tendency to associate with each other, or with any other similarly complementary overhang, by forming base pairs **(Fig.53)**.

Different fragments with the same sequences in their overhanging ends can be generated by:

1. cutting with the same restriction nuclease;
2. cutting with different restriction endonucleases that happen to recognize the same target sequence (isoschizomers); or
3. by cutting with enzymes which have different recognition sequences but happen to produce compatible sticky ends, for example BamHI and MboI.

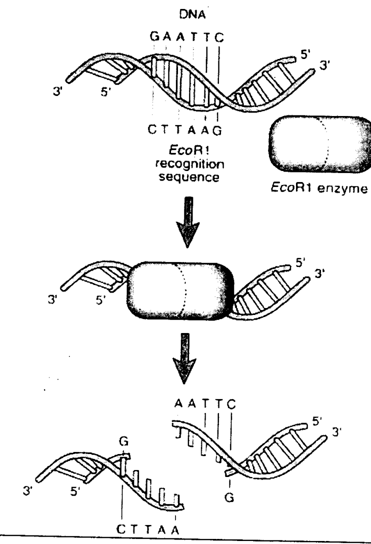


Figure 53. Restriction endonucleases generate overhanging ends

Cell-based DNA cloning

The recombinant DNA cloning involves following steps:

1. **Cutting:** This step is facilitated by cutting the target DNA and replicon molecules with the same specific restriction endonucleases

2. **Attachment**: In vitro covalent attachment (ligation) of the desired DNA fragments (target DNA) to a replicon depending on the presence of sticky ends in both target DNA and vector.

3. **Ligation:** Because the overhanging ends generated by restriction endonucleases are very short (typically four nucleotides or less), hydrogen bonding between complementary overhanging ends provides a rather weak contact between two molecules, and can only be maintained at low temperatures. However, it does facilitate subsequent covalent bonding between the two associated molecules (DNA ligation). This is performed using the enzyme DNA ligase. Ligation of blunt-ended fragments is also possible, although less efficient than sticky end ligation.

4. **Transformation**: The recombinant DNA molecules are transferred into host cells (often bacterial or yeast cells) in which the chosen replicon can undergo DNA replication independently of the host cell chromosome(s). The plasma membrane of cells is selectively permeable and does not normally admit large molecules such as long DNA fragments. However, cells can be treated in certain ways (e.g. by exposure to certain high ionic strength salts, short electric shocks, etc.) so that the permeability properties of the plasma membranes are altered. As a result, a fraction of the cells become competent, that is capable of taking up foreign DNA from the extracellular environment. Only a small percentage of the cells will take up the foreign DNA (DNA transformation).

5. **Screening:** Identification of cells containing the vector molecule requires engineering or selection of the vector molecule to contain a suitable marker gene whose expression provides a means of identifying cells containing it. Two popularly used marker gene systems are based on:

    1. **antibiotic resistance genes.** A host cell strain is chosen that is sensitive to a particular antibiotic, often ampicillin, tetracycline or chloramphenicol. The corresponding vector has been engineered to contain a gene which confers resistance to the antibiotic. After transformation, cells are plated on agar containing the antibiotic to rescue cells transformed by the vector.

    2. **b-galactosidase gene complementation.** The host cell is a mutant which contains a fragment of the b-galactosidase gene but cannot make any functional b-galactosidase. The vector is engineered to contain a different fragment of the b-galactosidase gene. After transformation by the vector, functional complementation occurs resulting in active b-galactosidase which can be assayed by acting on a colorless substance, Xgal (5-bromo, 4-chloro, 3-indolyl b-d-galactopyranoside), to make a blue product.

Recombinant screening

Identification of cells containing recombinant DNA (vector molecules with inserts) is often accomplished by insertional inactivation of the marker gene. The vector molecule is designed to have a multiple cloning site (polylinker) located within the marker gene. The number of nucleotides in the polylinker sequence containing the multiple cloning sites is a multiple of three, so that its insertion in the marker gene does not result in a shift in the translation reading frame. As a result of maintaining the reading frame, and the very small size of the polylinker, the activity of the marker gene is maintained despite the insertion of the polylinker.

However, if the vector then contains recombinant DNA inserted in the middle of the polylinker, the resulting large insertion causes loss of expression of the marker gene. In the case of a marker b-galactosidase gene, insertional inactivation means that cells containing recombinant DNA are colorless in the presence of Xgal, while cells containing nonrecombinant vector are blue.

6. **Selective propagation** of cell clones involves two stages. Initially the transformed cells are plated out by spreading on an agar surface in order to encourage the growth of well-separated cell colonies. These are cell clones (populations of identical cells all descended from a single cell). Subsequently, individual colonies can be picked from a plate and the cells can be further expanded in liquid culture. **(Fig. 54)**

Propagation

The transformed cells are allowed to multiply. In the case of cloning using plasmid vectors and a bacterial cell host, a solution containing the transformed cells is simply spread over the surface of nutrient agar in a petri dish. This usually results in the formation of bacterial colonies which consist of cell clones (identical progeny of a single ancestral cell). Picking an individual colony into a tube for subsequent growth in liquid culture permits a secondary expansion in the number of cells which can be scaled up to provide very large yields of cell clones. If the original cell contained a single type of foreign DNA fragment attached to a replicon, then so will the descendants, resulting in a huge amplification in the amount of the specific foreign fragment.

7. **Isolation of recombinant DNA clones** by harvesting expanded cell cultures and selectively isolating the recombinant DNA.

8. **Recovery:** Expanded cultures representing cell clones derived from a single cell can then be processed to recover the recombinant DNA. To do this, the cells are lysed, and the DNA is extracted and purified using procedures that result in recovery of the recombinant DNA.

Such procedures take advantage of differences between the recombinant DNA and the host chromosomal DNA. In the case of bacterial cells, the bacterial chromosome is circular double-stranded DNA, like any plasmids containing introduced foreign DNA. However, the chromosomal DNA is relatively very large (~4.6 Mb) compared with most recombinant DNA molecules

(often only a few kilobases long). During cell lysis and subsequent extraction procedures, the very large bacterial chromosomal DNA, but not the small recombinant plasmid DNA, will undergo nicking and shearing, generating linear DNA fragments with free ends. This difference can be exploited by subjecting the isolated DNA to a denaturation step, often as a result of exposure to alkaline pH in the 12.0-12.5 range. Following this treatment, the linearized host cell DNA readily denatures, but the strands of covalently closed circular (CCC) plasmid DNA are unable to separate.

After normal conditions have been restored, the two strands of the CCC DNA rapidly re-align in perfect register to form native superhelical molecules or so-called supercoiled DNA. This higher order form of twisting occurs in any CCC DNA because the tension introduced by twisting of the double helix cannot be relaxed, unlike in DNA with one or two free ends where relaxation is possible by rotation of a free end. The denatured host cell DNA precipitates out of solution, leaving behind the CCC plasmid DNA.

9. **Purification** If required, further purification is possible by column chromatography. Alternatively, equilibrium density gradient centrifugation (isopycnic centrifugation) is used: the partially purified DNA is centrifuged to equilibrium in a solution of cesium chloride containing ethidium bromide (EtBr). EtBr binds by intercalating between the base pairs, thereby causing the DNA helix to unwind. Unlike chromosomal DNA, a CCC plasmid DNA has no free ends and can only unwind to a limited extent, which limits the

amount of EtBr it can bind. EtBr-DNA complexes are denser when they contain less EtBr, so CCC plasmid DNA will band at a lower position in the cesium chloride gradient than chromosomal DNA and plasmid circles that are open, enabling separation of the recombinant DNA from host cell DNA. The resulting recombinant DNA molecules will normally be identical to each other (representing a single target DNA fragment) and are referred to as DNA clones **(Fig. 54)**.

Generally, ligation reactions are designed to promote the formation of recombinant DNA (by ligating target DNA to vector DNA), although vector cyclization, vector-vector concatemers and target DNA-target DNA ligation are also possible. To achieve this, the vector molecules are often treated so as to prevent or minimize their ability to undergo cyclization. There are two common ways of achieving this.

- **Cutting of vectors with two different restriction endonucleases**. It is often convenient to cut the vector with two restriction endonucleases which do not produce complementary overhanging ends (e.g. EcoRI and BamHI), and remove the small vector fragment between the sites, resulting in a vector molecule whose two ends cannot religate. However, if target DNA is cut with the same enzyme combination, recombinant DNA can easily be formed.

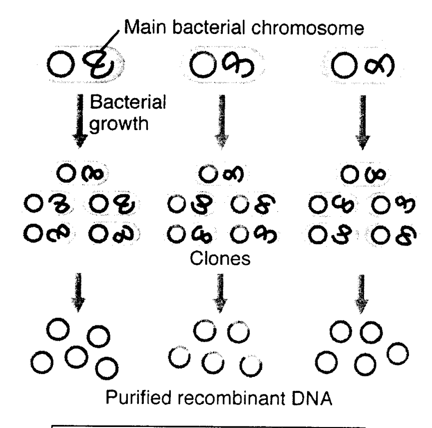


Figure 54. Propagation and purification

- **Vector dephosphorylation**: During DNA ligation in vitro, the enzyme DNA ligase will catalyze the formation of a phosphodiester bond only if one nucleotide contains a 5' phosphate group and the other contains a 3 hydroxyl group. If the 5' phosphate groups at both ends of the vector DNA are removed by treatment with alkaline phosphatase, the tendency for the vector DNA to recircularize will therefore be minimized. A foreign DNA insert can, however, provide 5-terminal phosphates which can then be joined to the 3' hydroxyl groups provided by the vector. This method, therefore,

increases the frequency of cells containing recombinant DNA.

Restriction mapping

Restriction mapping of DNA clones involves cutting the DNA with one or more of a series of different restriction endonucleases and separating the resulting fragments according to size by agarose gel electrophoresis. Because of the conspicuous deficiency in the CpG dinucleotide in vertebrate genomes, recognition sites that are GC-rich will occur comparatively less frequently than expected in vertebrate DNA, but will not be so rare in bacterial cell DNA **(Table 8).** Restriction mapping of a recombinant DNA clone immediately provides details of the length of the insert DNA and some information on the location of unique restriction sites which may be useful for sub-cloning purposes.

Detection of mutations by restriction mapping

Occasionally, a pathogenic mutation directly abolishes or creates a restriction site, enabling direct screening for the pathogenic mutation. For example, the sickle cell mutation is a single nucleotide substitution (A→ T) at codon 6 in the β-globin gene, which causes a missense mutation (Glu→ Val), and at the same time abolishes an MstII restriction site which spans codons 5 to 7. Consequently, a β -globin DNA probe can differentiate the normal A-globin and the mutant S-globin alleles in MstII-digested human DNA: the former exhibits 1.2 kb and 0.2 kb MstII fragments, whereas the sickle cell allele exhibits a 1.4 kb MstII fragment **(Fig. 55).**

Detection of conventional RFLPs

The great majority of mutations are not associated with disease; instead, they often occur within noncoding DNA sequences. As a large number of recognition sequences are known for type II restriction endonucleases, many point mutation polymorphisms will be characterized by alleles which possess or lack a recognition site for a specific restriction endonuclease and therefore display restriction site polymorphism (RSP). Accordingly, individual RSPs normally have two detectable alleles (one lacking and one possessing the specific restriction site). RSPs can be assayed by digesting genomic DNA samples with the relevant restriction endonuclease and identifying specific restriction fragments whose lengths are characteristic of the two alleles, so-called RFLPs.

Purposes of recombinant DNA technology

1. DNA cloning is geared simply towards amplifying the introduced DNA to obtain sufficient quantities for structural and functional studies.
2. To express the gene in some way (**expression cloning**). In each case, appropriate expression signals need to be provided by the cloning system.
3. To obtain an RNA product. Examples include generation of antisense RNA probes (**riboprobes**) for use in tissue in situ hybridization studies.

4. Generation of antisense RNA for inhibiting or destroying, expression of specific genes, either in functional studies or for therapeutic purposes.

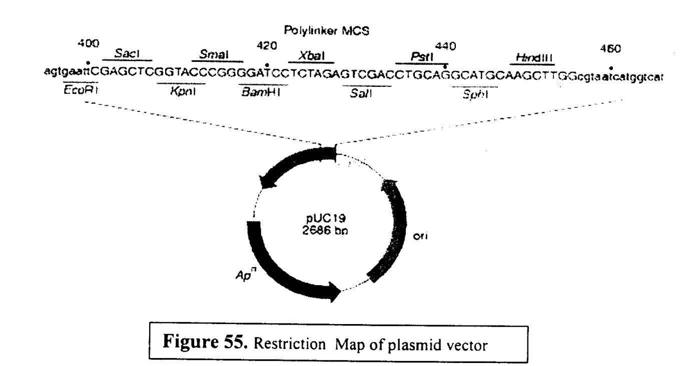


Figure 55. Restriction Map of plasmid vector

5. To retrieve large quantities of an expression product, as in the need to generate large quantities of a specific protein to assist subsequent crystallography studies or attempts to raise specific antibodies against the protein.

6. Expression of eukaryotic proteins in bacterial cells can also be used as a target system for screening with antibodies to identify previously uncharacterized genes.

7. Expression of important eukaryotic genes in bacterial cells can provide an endless bulk supply of important, medically relevant compounds or of proteins for basic follow-up

Table 8. Restriction endonucleases.

Enzyme	Source	Sequence cut	Average expected fragment size (kb) in human DNA[a]
AluI	Arthrobacter luteus	AGCT	0.3
HaeIII	Hemophilus aegyptus	GGCC	0.6
TaqI	Thermus aquaticus	TCGA	1.4
MnlI	Moraxella nonliquefaciens	CCTC/GAGG	0.4
HindIII	Hemophilus influenzae Rd	AAGCTT	3.1
EcoRI	Escherichia coli R factor	GAATTC	3.1
BamHI	Bacillus amyloliquefaciens H	GGATCC	7.0
PstI	Providencia stuartii	CTGCAG	7.0
MstI	Microcoleus species	CCTNAGG[c]	7.0
SmaI	Serratia marcescens	CCCGGG	78
BssHII	Bacillus stearothermophilu s	GCGCGC	390[b]
NotI	Norcadia otitidis-caviarum	GCGGCCGC	9766[b]

research studies. As the production of very large amounts of a protein (insulin hormone and growth hormone).

8. An important application of this approach is designed to permit the construction of an antibody specific for a eukaryotic polypeptide where none have previously been available (anti hepatitis B).

Summary

DNA technology are based on two different approaches:

1. **DNA cloning.** The desired fragment must be selectively amplified so that it is purified essentially to homogeneity.
2. **Molecular hybridization.** The fragment of interest is not amplified, but instead is specifically detected within a complex mixture of many different sequences.

DNA cloning

The desired DNA fragments selectively amplified, resulting in a programmed large increase in copy number of selected DNA sequences.:

- **Cell-free DNA cloning (PCR).**
- **Cell-based DNA cloning (recombinant DNA).** This is an in vivo cloning method. The system composed of.

1-Target DNA

2-Vectors (Replicons)

Cell cloning requires that the foreign DNA fragments which are introduced into a host cell must be able to replicate.

Replicons:

Could be provided in two ways.

1. **Chromosomal:** Introduce the fragments into the host cells chromosome and are propagated under the control of a host cell replicon. (retroviruses)
2. **Extra-chromosomal replicon:** The foreign DNA fragments are attached in vitro to a purified replicon and the resulting hybrid molecules are transferred into host cells, where they replicate independently of the host cell chromosomes.

Extra-chromosomal replicons (vectors)

Basic types of extra-chromosomal replicons are:

1. **Plasmids** are small circular double-stranded DNA molecules which individually contain very few genes.
2. **Bacteriophages** are viruses which infect bacterial cells.
3. **Cosmid vectors:** Cosmid vectors contain cos sequences inserted into a small plasmid vector.

4. **BACs (Bacterial artificial chromosome) vectors:** (BACs contain a low copy number replicon, .
5. **PACs(P1-derived artificial chromosome)**: bacteriophage packaging systems with P1 vectors
6. **YACs (yeast artificial chromosomes):** the main advantage offered by YACs has been the ability to clone very large DNA fragments.
7. **Phagemid** are plasmids which have been artificially manipulated so with phage,

Host cells

Bacterial cells are widely used because of their capacity for rapid cell division..

- **Restriction endonucleases**

which cleave DNA whenever a small, specific recognition sequence, usually 48 base pairs (bp) long occurs. The recognition sequences for the vast majority of type II restriction endonucleases are normally palindromes, that is the sequence of bases is the same on both strands when read in the 5'→3' direction.

Cell-based DNA cloning

The recombinant DNA cloning involves following steps:

1. **Cutting**: The target DNA and replicon molecules with the same specific restriction endonucleases
2. **Attachment:** In vitro covalent attachment (ligation) of the desired DNA fragments (target DNA) to a replicon.
3. **Ligation** between the two associated molecules (DNA ligation) using DNA ligase.
4. **Transformation**: The recombinant DNA molecules are transferred into host cells. The cells will take up the foreign DNA (DNA transformation).
5. **Screening:** Identification of cells containing the vector molecule based on:
    1. **antibiotic resistance genes.**
    2. **b-galactosidase gene complementation.**
6. **Selective propagation** of cell clones .
7. **Isolation of recombinant DNA clones** by harvesting expanded cell cultures and selectively isolating the recombinant DNA.

8. **Recovery:** Expanded cultures representing cell clones derived from a single cell can then be processed to recover the recombinant DNA.
9. **Purification:** If required, further purification is possible by column chromatography.

Restriction mapping

Restriction mapping of DNA clones involves cutting the DNA with one or more of a series of different restriction endonucleases and separating the resulting fragments according to size by agarose gel electrophoresis..

Detection of mutations by restriction mapping

Occasionally, a pathogenic mutation directly abolishes or creates a restriction site, enabling direct screening for the pathogenic mutation.

Purposes of recombinant DNA technology

1. DNA cloning for structural and functional studies.
2. To express the gene in some way (**expression cloning**).
3. To obtain an RNA product.
4. Generation of antisense RNA for inhibiting or destroying expression of specific genes, either in functional studies or for therapeutic purposes.
5. To retrieve large quantities of an expression product.
6. Identification of uncharacterized genes.
7. Supply of important, medically relevant compounds or of proteins (insulin hormone and growth hormone).
8. Antibody specific for a eukaryotic polypeptide where none have previously been available (anti hepatitis B).

CHAPTER 11



PCR

Introduction

The polymerase chain reaction (PCR) has revolutionized molecular genetics by permitting rapid cloning and analysis of DNA. Since the first reports describing this new technology in the mid 1980s, there have been numerous applications in both basic and clinical research. Two other fundamental technologies are DNA sequencing and in vitro mutagenesis, both of which can be accomplished using PCR-based and non PCR-based methods.

Sstandard PCR reaction

PCR is a rapid and versatile in vitro method for amplifying defined target DNA sequences present within a source of DNA. Usually, the method is designed to permit selective amplification of a specific target DNA sequence(s) within a heterogeneous collection of DNA sequences (e.g. total genomic DNA or a complex cDNA population).

To permit such selective amplification, some prior DNA sequence information from the target sequences is required. This information is used to design two oligonucleotide primers (amplimers) which are specific for the target sequence and which are often about 15-25 nucleotides long.

After the primers are added to denatured template DNA, they bind specifically to complementary DNA sequences at the target site. In the presence of a suitably heat-stable DNA polymerase and DNA precursors (the four deoxynucleoside

triphosphates, dATP, dCTP, dGTP and dTTP), they initiate the synthesis of new DNA strands which are complementary to the individual DNA strands of the target DNA segment, and which will overlap each other.

The PCR is a chain reaction because newly synthesized DNA strands will act as templates for further DNA synthesis in subsequent cycles. After about 25 cycles of DNA synthesis, the products of the PCR will include, in addition to the starting DNA, about 10^5 copies of the specific target sequence, an amount which is easily visualized as a discrete band of a specific size when submitted to agarose gel electrophoresis. A heat-stable DNA polymerase is used because the reaction involves sequential cycles composed of three steps:

- **Step1**: The solution is heated to 95°C to dissociate the two chains of the double-stranded target DNA **(Fig.56-A);**
- **Step 2**: The solution is then cooled to about 55°C to allow the primers to bind (anneal) to the ends of the DNA strands **(Fig.56-B)**;
- **Step 3**: The solution is reheated to about 75°C, the optimal temperature for the Taq polymerase to synthesize complementary copies of each strand. One PCR cycle takes less than two minutes to complete **(Fig.56-C).**

Discovery of the Taq polymerase,(Suitably heat-stable DNA polymerases) have been obtained from microorganisms which lives in the hot springs of Yellowstone National Park and can survive heating to 95°C, was a key to the immense success of PCR because it meant that the enzyme could be added just once to the reaction tube. Other polymerases were destroyed at 95°C,

so new enzyme had to be added at each cycle. Eliminating the need to open the tubes between each cycle to add new

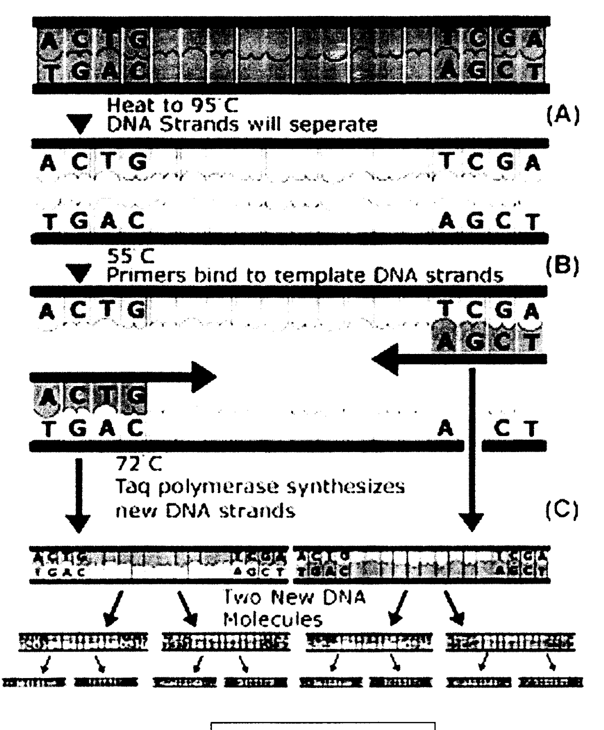


Figure 56. Steps of PCR

polymerase reduces the chance of contamination and increases the reliability and repeatability of the procedure.

Primer design

The specificity of amplification depends on the extent to which the primers can recognize and bind to sequences other than the intended target DNA sequences. For complex DNA sources, such as total genomic DNA from a mammalian cell, it is often sufficient to design two primers about 20 nucieotides long. This is because the chance of an accidental perfect match elsewhere in the genome for either one of the primers is extremely low, and for both sequences to occur by chance in close proximity in the specified direction is normally exceedingly low.

1. **Nested primers**: The products of an initial amplification reaction are diluted and used as the target DNA source for a second reaction in which a different set of primers is used, corresponding to sequences located close, but internal, to those used in the first reaction.
2. **Hot-start PCR**: Mixing of all PCR reagents prior to an initial heat denaturation step allows more opportunity for nonspecific binding of primer sequences. To reduce this possibility, one or more components of the PCR are physically separated until the first denaturation step. A popular approach is to use a specially formulated wax bead designed to fit snugly within a PCR reaction tube. The reaction components minus the enzyme and reaction buffer are added to the tube followed

by the molten wax bead which floats on top and then solidifies on cooling. The thermostable polymerase is then added with buffer. At the initial denaturation step the wax melts again and rises to the surface causing all the reaction components to come into contact with each other.

3. **Touch-down PCR**: Most thermal cyclers can be programmed to perform runs in which the annealing temperature is lowered during the PCR cycling from an initial value above the expected T_m to a value below the T_m. By keeping the stringency of hybridization initially very high, the formation of spurious products is discouraged, allowing the expected sequence to predominate.

Advantages of PCR

Because of its simplicity, PCR is a popular technique with a wide range of applications which depend on essentially three major advantages of the method.

1-Speed and ease of use

DNA cloning by PCR can be performed in a few hours, using relatively unsophisticated equipment. Typically, a PCR reaction consists of 30 cycles containing a denaturation, synthesis and reannealing step, with an individual cycle typically taking 35 min in an automated thermal cycler. This compares favorably with the time required for cell-based DNA cloning, which may take weeks. Clearly, some time is also required for designing and

synthesizing oligonucleotide primers, but this has been simplified by the availability of computer software for primer design and rapid commercial synthesis of custom oligonucleotides. Once the conditions for a reaction have been tested, the reaction can then be repeated simply.

2-Sensitivity

PCR is capable of amplifying sequences from minute amounts of target DNA, even the DNA from a single cell. Such exquisite sensitivity has afforded new methods of studying molecular pathogenesis and has found numerous applications in forensic science, in diagnosis, in genetic linkage analysis using single-sperm typing and in molecular paleontology studies, where samples may contain minute numbers of cells. However, the extreme sensitivity of the method means that great care has to be taken to avoid contamination of the sample under investigation by external DNA, such as from minute amounts of cells from the operator.

3-Robustness

PCR can permit amplification of specific sequences from material in which the DNA is badly degraded or embedded in a medium from which conventional DNA isolation is problematic. As a result, it is again very suitable for molecular anthropology and paleontology studies, for example the analysis of DNA recovered from archaeological remains. It has also been used successfully to amplify DNA from formalin-fixed tissue samples, which has important applications in molecular pathology and, in some cases, genetic linkage studies.

Disadvantages of PCR

Despite its huge popularity, PCR has certain limitations as a method for selectively cloning specific DNA sequences.

1-Need for target DNA sequence

In order to construct specific oligonucleotide primers that permit selective amplification of a particular DNA sequence, some prior sequence information is necessary. This normally means that the DNA region of interest has been partly characterized previously, often following cell-based DNA cloning. However, a variety of techniques have been developed that exclude the need for prior DNA sequence information of the target DNA.

2-Short size PCR product

A clear disadvantage of PCR as a DNA cloning method has been the size range of the DNA sequences that can be cloned. Unlike cell-based DNA cloning where the size of cloned DNA sequences can approach 2 Mb, reported DNA sequences cloned by PCR have typically been in the 0.15 kb size range.

Recently, however, conditions have been identified for effective amplification of longer targets, including a 42-kb product from the bacteriophage l genome. Often, the conditions for long range PCR involve a combination of modifications to standard conditions with a two-polymerase system.

The amount of PCR product obtained in a single reaction is also much more limited than the amount that can be obtained using cell-based cloning where scale-up of the volumes of cell cultures is possible. The efficiency of a PCR reaction will vary from template to template and according to various factors that

are required to optimize the reaction but typically only comparatively small amounts of product are achieved[1]

3-Infidelity of DNA replication

Cell-based DNA cloning involves DNA replication in vivo, which is associated with a very high fidelity of copying because of proofreading mechanisms. However, when DNA is replicated in vitro the copying error rate is considerably greater. Taq DNA polymerase does not confer a proofreading function, and the error rate due to base misincorporation during DNA replication is rather high: for a 1 kb sequence that has undergone 20 effective cycles of duplication, approximately 40% of the new DNA strands synthesized by PCR using this enzyme will contain an incorrect nucleotide resulting from a copying error. This means that, even if the PCR reaction involves amplification of a single DNA sequence, the final product will be a mixture of extremely similar, but not identical DNA sequences.

Despite the errors due to replication in vitro, DNA sequencing of the total PCR product may give the correct sequence. This is because, although individual DNA strands in the PCR product often contain incorrect bases, the incorporation of incorrect bases is essentially random. As a result, for each base position, the contribution of one incorrect base on one or more strands is overwhelmed by the contributions from the huge majority of strands which will have the correct sequence.

More recently, the problem of infidelity of DNA replication during the PCR reaction has been considerably reduced by using alternative heat-stable DNA polymerases which have associated 3' 5' exonuclease activity. For example, the Pyrococcus furiosus DNA polymerase is becoming more widely

used because of the proofreading conferred by its associated 3' 5' exonuclease activity. The resulting PCR product has a much lower level of mutations introduced by copying errors: for a 1 kb segment of DNA that has undergone 20 effective cycles of duplication, about 3.5% of the DNA strands in the product carry an altered base.

4-Limited amount of PCR product

The amount of material that can be cloned in a single PCR reaction is limited, and it is time-consuming and expensive to repeat the same PCR reaction many times to achieve large quantities of the desired DNA. In addition, the PCR product may not be in a suitable form that will permit some subsequent studies. As a result, it is often convenient to clone the PCR product in a cell-based cloning system in order to obtain large quantities of the desired DNA and to permit a variety of analyses. Various plasmid cloning systems are used to propagate PCR-cloned DNA.

Applications of PCR

Although PCR was first developed only a decade and a half ago, the simplicity and the versatility of the technique have ensured that it is among the most ubiquitous of molecular genetic methodologies, with a wide range of general applications.

1-DNA labeling by PCR

The standard PCR reaction can be modified to permit incorporation of labeled nucleotides. Two methods are commonly used:

- **Standard PCR-based DNA labeling**: The PCR reaction is modified to include one or more labeled nucleotide precursors which become incorporated into the PCR product throughout its length.

- **Primer-mediated 5'end labeling**: PCR is conducted using a primer in which a labeled group is attached to the 5 end. As PCR proceeds the primer with its 5 end-label is incorporated into the PCR product. This method is often used with fluorophore labels during DNA sequencing and is an example of a general PCR mutagenesis method known as **5add-on mutagenesis** which has many applications.

2-Mutations screening

Because of its rapidity and simplicity, PCR is ideally suited to providing numerous DNA templates for mutation screening. Partial DNA sequences, at the genomic or the cDNA level, from a gene associated with disease, or some other interesting phenotype, immediately enable gene-specific PCR reactions to be designed. Amplification of the appropriate gene segment then enables rapid testing for the presence of associated mutations in large numbers of individuals. By contrast, cell-based DNA cloning of the gene from numerous different individuals is far too slow and labor-intensive to be considered as a serious alternative.

Typically, the identification of exon-intron boundaries and sequencing of the ends of introns of a gene of interest offers the possibility of genomic mutation screening. Individual exon-specific amplification reactions are developed by designing primers which recognize intronic sequences located close to the

exon-intron boundary. The resulting PCR products are then analyzed by rapid mutation-screening methods.

3-Genotyping for polymorphic markers

Restriction site polymorphisms (RSPs) result in lacking a specific restriction site.. As a convenient, PCR can type RSPs by simply designing primers using sequences which flank the polymorphic restriction site, amplifying from genomic DNA, then cutting the PCR product with the appropriate restriction enzyme and separating the fragments by agarose gel electrophoresis.

4-Allele-specific PCR (ARMS test)

Oligonucleotide primers can be designed so as to discriminate between target DNA sequences that differ by a single nucleotide in the region of interest. This is a form of allele-specific PCR, the PCR equivalent of the allele-specific hybridization.

5-DOP-PCR (degenerate oligonucleotide-primed PCR)

DOP-PCR is a form of PCR which is deliberately designed to permit possible amplification of several products. The two primers may be partially degenerate oligonucleotides, composed of panels of oligonucleotide sequences that have the same base at certain nucleotide positions, but are different at others. As a result, there may be comparatively many primer binding sites in the source DNA. This provides a means of searching for a new or uncharacterized DNA sequence that belongs to a family of related sequences either within or between species.

6-Linker-primed PCR (ligation adaptor PCR)

Another way of enabling amplification of essentially all DNA sequences in a complex DNA mixture involves first ligating a known sequence to all fragments. To do this, the target DNA population is digested with a suitable restriction endonuclease, and double-stranded oligonucleotide linkers (also called **adaptors**) with a suitable overhanging end are ligated to the ends of target DNA fragments. Amplification is then performed using oligonucleotide primers which are specific for the linker sequences. In this way, all fragments of the DNA source which are flanked by linker oligonucleotides can be amplified.

7-Anchored PCR

It is often desirable to be able to amplify previously uncharacterized DNA sequences that neighbor a known DNA sequence, either at the genomic or cDNA level. To do this a form of anchored PCR is used. One of the primers is specific for the target sequence and the second primer is specific for a common.

8-Reverse transcriptase-PCR

Such cDNA mutation screening may be the only way in which mutations can be screened if the exon-intron organization of a gene has not been established. To do this, mRNA is isolated from a convenient source of tissue, such as blood cells, converted into cDNA using reverse transcriptase and the cDNA is used as a template for a PCR reaction. This version of the standard genomic PCR reaction is consequently often referred to as RT-PCR (reverse transcriptase-PCR). Clearly, the method is ideally suited to genes expressed at high levels in easily accessible cells, such as blood cells and liver cells.

9- DNA sequencing

Formerly, chemical DNA sequencing methods were often employed, using base-specific chemical modification and subsequent cleavage of the DNA. Currently, however, the vast majority of DNA sequencing is carried out using an enzymatic method: the DNA to be sequenced is provided in a single-stranded form from which DNA polymerase synthesizes new complementary DNA strands. Usually, the single-stranded DNA template is obtained using a cloning system which permits recovery of single-stranded recombinant DNA, as with M13 or phagemid cloning systems. The subsequent DNA sequencing reactions involve DNA synthesis using one or more labeled nucleotides and a universal sequencing primer that is complementary to the vector sequence flanking the cloning site.

Cycle sequencing by PCR

In addition to the normal nucleotide precursors, DNA synthesis is carried out in the presence of base-specific dideoxynucleotides (**ddNTPs**). The latter are analogs of the normal dNTPs but differ in that they lack a hydroxyl group at the 3 carbon position as well as the 2 carbon. A dideoxynucleotide can be incorporated into the growing DNA chain by forming a phosphodiester bond between its 5 carbon atom and the 3 carbon of the previously incorporated nucleotide. However, since ddNTPs lack a 3 hydroxyl group, any ddNTP that is incorporated into a growing DNA chain cannot participate in phosphodiester bonding at its 3 carbon atom, thereby causing abrupt termination of chain synthesis.

Four parallel base-specific reactions are conducted using a mix of all four dNTPs and also a small proportion of one of the four ddNTPs. By setting the concentration of the ddNTP to be

very much lower than that of its normal dNTP analog, chain termination will occur randomly at one of the many positions containing the base in question. Each reaction is therefore a partial reaction: chain termination occurs randomly at one of the possible bases in any one DNA strand. However, the DNA to be sequenced in a DNA sequencing reaction is a population of (usually) identical molecules. As a result, each one of the four base-specific reactions will generate a collection of labeled DNA fragments of different sizes, with a common 5end but variable 3 ends (the common 5 end is defined by the sequencing primer and the 3 ends which terminate with the chosen ddNTP are variable because the insertion of the dideoxynucleotide occurs randomly at one of the many different positions that will accept that specific base.

Fragments that differ in size by even a single nucleotide can be separated on a denaturing polyacrylamide gel **(Fig.57)**. The differently sized fragments can be detected by incorporating labeled groups into the reaction products, either by incorporating labeled nucleotides or by using a primer with a labeled group. The sequence can then be read off by reading from the bottom of the gel to the top, a direction that gives the 5 3 sequence of the complementary strand of the provided DNA template.

Sequencing by automated detection systems

Traditional dideoxy sequencing methods have employed radioisotope labeling: the dNTP mix contains a proportion of radiolabeled nucleotides which are incorporated within the growing DNA chains. Following electrophoresis, the gel is dried and an autoradiographic film is placed in contact with the dried gel. After a suitable exposure time, the film is developed, giving

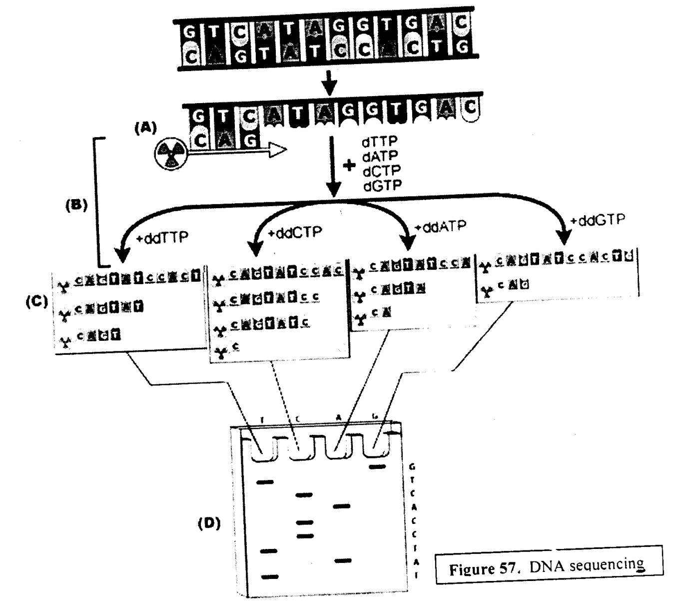


Figure 57. DNA sequencing

a characteristic pattern of dark bands. ^{32}P-labeled nucleotides are not very suitable for this purpose: the high energy b-radiation causes considerable scattering of the signal, leading to diffuse bands. Instead, ^{35}S- or ^{33}P-labeled nucleotides have been used.

Large-scale DNA sequencing efforts are dependent on improving efficiency by partial automation of the technologies involved. One major improvement in recent years has been the

development of automated procedures for fluorescent DNA sequencing. These procedures generally use primers or dideoxynucleotides to which are attached fluorophores (chemical groups capable of fluorescing. During electrophoresis, a monitor detects and records the fluorescence signal as the DNA passes through a fixed point in the gel. The use of different fluorophores in the four base-specific reactions means that, unlike conventional DNA sequencing, all four reactions can be loaded into a single lane. The output is in the form of intensity profiles for each of the differently colored fluorophores, but the information is simultaneously stored electronically. This precludes transcription errors when an interpreted sequence is typed by hand into a computer file. Recent advances in technology mean that the accuracy of DNA sequencing using automated methods is acceptably high.

DNA microarray technology

DNA sequencing can be accomplished by hybridization of the target DNA to a series of oligonucleotides of known sequence, usually about 78 nucleotides long. If the hybridization conditions are specific, it is possible to check which oligonucleotides are positive by hybridization, feed the results into a computer and use a program to look for sequence overlaps in order to establish the required DNA sequence. DNA microarrays have permitted sequencing by hybridization to oligonucleotides on a large scale and in a test system, the sequence of human mtDNA previously first determined in 1981 was recently re-sequenced by DNA microarray hybridization. This type of technology is increasing in importance for assessing sequence variation over at least modest lengths of DNA and diagnostic applications in mutation analysis are proliferating.

١٥-In vitro site-specific mutagenesis

Mutagenesis is a fundamentally important DNA technology which seeks to change the base sequence of DNA and test its effect on gene or DNA function. The mutagenesis can be conducted in vivo (in studies of model organisms, or cultured cells) or in vitro and the mutagenesis can be directed to a specific site in a pre-determined way (site-directed mutagenesis), or can be random. In the case of in vivo mutagenesis, for example, gene targeting offers exquisite site-directed mutagenesis within living cells while exposure of male mice to high levels of a powerful mutagen such as ethyl nitrosurea (ENU) and subsequent mating of the mice offers a form of random mutagenesis which can be important in generating new mutants.

In vitro mutagenesis can involve essentially random approaches to mutagenesis, which may be valuable in producing libraries of new mutants. In addition, if a gene has been cloned and a functional assay of the product is available, it is also very useful to be able to employ a form of in vitro mutagenesis which results in alteration of a specific amino acid or small component of the gene product in a predetermined way.

Oligonucleotide mismatch mutagenesis is a popular method of introducing a predetermined single nucleotide change into a cloned gene.

Many in vitro assays of gene function wish to gain information on the importance of individual amino acids in the encoded polypeptide. This may be relevant when attempting to assess whether a particular missense mutation found in a known disease gene is pathogenic, or just generally in trying to evaluate the contribution of a specific amino acid to the biological

function of a protein. A popular general approach involves cloning the gene or cDNA into an M13 or phagemid vector which permits recovery of single-stranded recombinant DNA. A mutagenic oligonucleotide primer is then designed whose sequence is perfectly complementary to the gene sequence in the region to be mutated, but with a single difference: at the intended mutation site it bears a base that is complementary to the desired mutant nucleotide rather than the original. The mutagenic oligonucleotide is then allowed to prime new DNA synthesis to create a complementary full-length sequence containing the desired mutation. The newly formed heteroduplex is used to transform cells, and the desired mutant genes can be identified by screening for the mutation.

Mismatched primer mutagenesis The primer is designed to be only partially complementary to the target site but in such a way that it will still bind specifically to the target. Inevitably this means that the mutation is introduced close to the extreme end of the PCR product. As described this approach may be exploited to introduce an artificial diagnostic restriction site that permits screening for a known mutation. Mutations can also be introduced at any point within a chosen sequence using mismatched primers. Two mutagenic reactions are designed in which the two separate PCR products have partially overlapping sequences containing the mutation. The denatured products are combined to generate a larger product with the mutation in a more central location.

Electrophoresis

Electrophoresis is used to separate pieces of DNA that vary in length. DNA is essentially sifted through a thin slab of "gel",

similar in consistency to over-hardened jello. The gel can be made of different materials - agarose or acrylamide - depending on the size of the DNA fragments to be separated. Liquid gel is poured into a mold leaving small slits or wells in one end where DNA samples (suspended in a buffer and dye) are placed once the gel solidifies. This illustration shows loading of a DNA sample in lanes 1 & 2, and DNA already run through the gel in lanes 3 & 4. In reality, all lanes are run together with a control sample in one lane (lane 5) containing a "ladder" of DNA fragments of known size to help estimate the sample size **(Fig. 57)**.

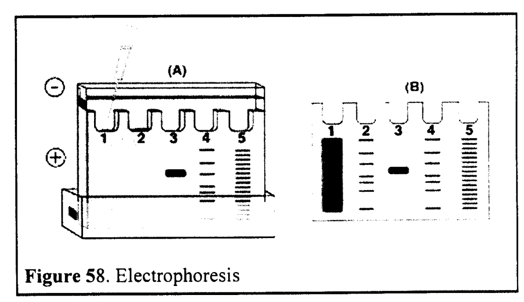


Figure 58. Electrophoresis

Once the DNA is loaded, the gel is immersed in a buffer solution and placed under a continuous electrical current. The DNA molecule has an overall negative electrical charge due to its chemical structure, so it is attracted to the positive end of the gel and slowly moves through the gel matrix **(Fig.58 A).** Typically, gels are run for several hours. After the run, DNA trapped in the gel can be visualized using the chemical **ethidium bromide**, which integrates into the groves of DNA, and fluoresces under

ultraviolet light. DNA may appear as a smear down each lane **(Fig.58 B** lane 1), or in the case of a PCR sample, as a discrete heavy band since most of the DNA is of one length **(Fig.58 B** lane 3).

11 -DNA Fingerprinting

DNA Fingerprinting is actually a patented process, but the term has been adopted to describe the analysis of repetitive sequences DNA fingerprinting takes advantage of the natural presence of repetitive regions of DNA sequence within the genome. These regions of DNA do not contain genes, that is, they are non-coding. Some of these repetitive regions are believed to play a role in maintaining chromosome structure, recombination, and/or regulatory control. Because these regions are non-coding, they accumulate mutations faster than coding regions, where mutations are much more likely to affect the individual's survival, and thus tend not be transmitted to the next generation. Since mutations occur randomly, each individual carries a unique set of these repetitive sequences.

DNA fingerprinting is performed either by probing DNA with markers that contain the repetitive sequences, (see before) or by using PCR to amplify specific repeat regions within the genome. Two types of DNA fingerprinting techniques are described below:

- **Variable Number Tandem Repeat** (VNTR) analysis relies on regions of DNA that contain different numbers of short, repeating sequences in different individuals, and at different positions in the genome (called loci). Genomic

DNA is fragmented at restriction sites that flank the VNTRs. Analysis of the digested DNA by Southern blot reveals a unique pattern of bands based on the number of repeats in the individual. Whereas one person may have two repeats at each of three loci, another may have one copy at one locus and two at both of the others, thus the pattern will vary at the third location. Depending on the length of the repeat, VNTRs are classified as microsatellite (12 - 200 bases) or minisatellite (2 - 4 bases).

- PCR-based methods of fingerprinting identify unique profiles of DNA fragments by varying the nature of PCR primer and the conditions under which the primer anneals. The conditions can be varied so that only sequences with exact complementarity to the primer sequence will bind; or at the other extreme, so that all sequences that are somewhat similar will bind. A profile of different sizes of bands is generated that reflects the various loci targeted by the primer. PCR-based fingerprinting is typically used to screen whole genomes, producing a large number of individual-specific bands. PCR-based methods include, randomly amplified polymorphic DNA (RAPD), arbitrarily primed PCR (AP-PCR), and amplified length fragment polymorphism (AFLP).

Summary

The polymerase chain reaction (PCR) has revolutionized molecular genetics by permitting rapid cloning and analysis of DNA..

Standard PCR reaction

To permit such selective amplification, we need (1) target DNA segment, (2) two oligonucleotide primers (amplimers), (3) DNA polymerase (Taq) and (4) nucleotide precursors.

The reaction involves sequential cycles composed of three steps:

- **Step1**: denaturation by heating to 95°C
- **Step2: annealing. to allow the primers to bind (anneal)**
- **Step3**:Taq polymerase synthesize complementary copies

Advantages of PCR. PCR has many advantage:

1-Speed and ease of use

DNA cloning by PCR can be performed in a 30 minutes.

2-Sensitivity

PCR is capable of amplifying sequences from minute amounts of target DNA, even the DNA from a single cell.

3-Robustness

PCR can permit amplification of specific sequences from material in which the DNA is badly degraded.

Disadvantages of PCR Despite its huge popularity, PCR has certain limitations as a method for selectively cloning specific DNA sequences.

1-Need for target DNA sequence

In order to construct specific oligonucleotide primers that permit selective amplification of a particular DNA sequence.

2-Short size PCR product

A clear disadvantage of PCR as a DNA cloning method is the small size range of the DNA sequences that can be cloned.

3-The amount of PCR product

Much limited than the amount that can be obtained using cell-based cloning.

4-Infidelity of DNA replication

Taq DNA polymerase does not confer a proofreading function, and the error rate due to base misincorporation during DNA replication is rather high: DNA strands synthesized by PCR using this enzyme will contain an incorrect nucleotide resulting from a copying error.

Applications of PCR

1-DNA labeling

The standard PCR reaction can be modified to permit incorporation of labeled nucleotides.

2-Mutations screening

PCR is ideally suited to providing numerous DNA templates for mutation screening.

3-Genotyping for polymorphic markers.

4-Allele-specific PCR (ARMS test)

Oligonucleotide primers can be designed so as to discriminate between target DNA sequences that differ by a single nucleotide.

5-DOP-PCR (degenerate oligonucleotide-primed PCR)

DOP-PCR is a form of PCR which is deliberately designed to permit possible amplification of several products.

6-Linker-primed PCR (ligation adaptor PCR)

DNA mixture is linked to a known sequence (**adaptors**) Amplification is performed using oligonucleotide primers which are specific for the linker sequences. so, all fragments of the DNA source which are flanked by linker oligonucleotides can be amplified.

7-Anchored PCR

One of the primers is specific for the target sequence and the second primer is specific for a common sequence.

8-Reverse transcriptase-PCR

mRNA is isolated from a tissue, converted into cDNA using reverse transcriptase and the cDNA is used as a template for a PCR reaction.

9-DNA sequencing

The DNA sequencing reactions involve DNA synthesis using one or more labeled nucleotides and a universal sequencing primer. DNA synthesis is carried out in the presence of base-specific dideoxynucleotides (**ddNTPs**) (chain termination molecule).

10-In vitro site-specific mutagenesis

PCR can be used to couple desired sequences or chemical groups to a target sequence and to produce specific pre-determined mutations in DNA sequences.

DNA microarray technology

DNA microarrays have permitted sequencing by hybridization to oligonucleotides on a large scale and in a test system, This type of technology is increasing in importance for assessing sequence variation over at least modest lengths of DNA and diagnostic applications in mutation analysis are proliferating.

Electrophoresis

Electrophoresis is used to separate pieces of DNA that vary in length. DNA is essentially sifted through a thin slab of "gel", similar in consistency to over-hardened jello. The gel can be made of different materials - agarose or acrylamide - depending on the size of the DNA fragments to be separated.

٤١-DNA fingerprinting

DNA fingerprinting takes advantage of the natural presence of repetitive regions of DNA sequence within the genome. Each individual carries a unique set of these repetitive sequences. DNA fingerprinting is performed either by probing DNA with markers that contain the repetitive sequences, or by using PCR to amplify specific repeat regions within the genome.

CHAPTER 12

NUCLEIC ACID HYBRIDIZATION

Introduction

Nucleic acid hybridization is a fundamental tool in molecular genetics which takes advantage of the ability of individual single-stranded nucleic acid molecules to form double-stranded molecules (that is, to hybridize to each other). For this to happen, the interacting single-stranded molecules must have a sufficiently high degree of base complementarity. Standard nucleic acid hybridization assays involve using a labeled nucleic acid probe to identify related DNA or RNA molecules (that is, ones with a significantly high degree of sequence similarity) within a complex mixture of unlabeled nucleic acid molecules, the target nucleic acid.

Numerous applications in molecular genetics involve taking an individual DNA clone and using it as a hybridization probe to screen for the presence of related sequences within a complex target of uncloned DNA or RNA. Sometimes the assay is restricted to simply checking for presence or absence of sequences related to the probe. In other cases, useful information can be obtained regarding the size of the complementary sequences, their subchromosomal location or their locations within specific tissues or groups of cells.

Preparation of nucleic acid probes

In standard nucleic acid hybridization assays the probe is labeled in some way. Nucleic acid probes may be made as single-

stranded or double-stranded molecules, but the working probe must be in the form of single strands.

DNA probes: Conventional DNA probes are isolated by

1- Cell-based DNA cloning: the starting DNA range in size from 0.1 kb to hundreds of kilobases in length and is usually originally double-stranded.

2- PCR: PCR-derived DNA probes have often been less than 10 kb long and are, originally double-stranded. Conventional DNA probes are labeled by incorporating labeled dNTPs during an *in vitro* DNA synthesis reaction.

RNA probes can conveniently be generated from DNA which has been cloned in a specialized plasmid vector. Such vectors normally contain a phage promoter sequence immediately adjacent to the multiple cloning site. An RNA synthesis reaction is employed using the relevant phage RNA polymerase and the four rNTPs, at least one of which is labeled. Specific labeled RNA transcripts can then be generated from the cloned insert.

Oligonucleotide probes are short (typically 15 - 50 nucleotides) single-stranded pieces of DNA made by chemical synthesis: mononucleotides are added, one at a time, to a starting mononucleotide.

Isotopic labeling and detection

Traditionally, labeling of nucleic acids has been conducted by incorporating nucleotides containing radioisotopes. Such radiolabeled probes contain nucleotides with a radioisotope

(often ^{32}P, ^{33}P, ^{35}S or ^{3}H), which can be detected specifically in solution or, much more commonly, within a solid specimen (autoradiography).

The intensity of an autoradiographic signal is dependent on the intensity of the radiation emitted by the radioisotope, and the time of exposure, which may often be long (one or more days, or even weeks in some applications). ^{32}P has been used widely in Southern blot hybridization, dot-blot hybridization, colony and plaque hybridization (see below) because it emits high energy b-particles which afford a high degree of sensitivity of detection. It has the disadvantage, however, that it is relatively unstable

Nonisotopic labeling and detection

Non-isotopic labeling systems involve the use of non-radioactive probes. Although developed only comparatively recently, they are becoming increasingly popular and are finding increasing applications in a variety of different areas. Fluorescence labeling is a popular nonisotopic labeling approach and detection is accomplished by fluorescence microscopy.

Dot-blot hybridization

The general procedure of dot-blotting involves taking an aqueous solution of target DNA, for example total human genomic DNA, and simply spotting it on a nitrocellulose or nylon membrane then allowing it to dry. The target DNA sequences are denatured, either by exposure to heat, or to alkali. The denatured target DNA sequences on the membrane are exposed to a solution containing single-stranded labeled probe sequences. After allowing sufficient time for probe-target heteroduplex formation,

the probe solution is decanted, and the membrane is washed to remove excess probe. It is then dried and exposed to an autoradiographic film **(Fig. 59)**.

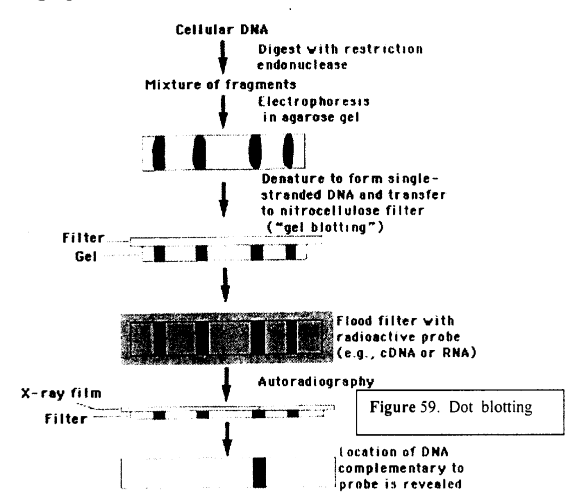


Figure 59. Dot blotting

Allele-specific oligonucleotide (ASO) probes is a useful application of dot-blotting which distinguish between alleles that differ in a single nucleotide. ASO probes are typically 15 - 20 nucleotides long and are normally employed under hybridization conditions at which the DNA duplex between probe and target is stable only if there is perfect base complementarity between them (a single mismatch between probe and target sequence is

sufficient to render the short heteroduplex unstable). Although ASOs can be used in conventional Southern blot hybridization, it is more convenient to use them in dot-blot assays.

Reverse dot-blotting. Another method of ASO dot blotting uses a reverse dot-blotting approach. This means that the oligonucleotide probes are not labeled and are fixed on a filter or membrane whereas the target DNA is labeled and provided in solution. Positive binding of labeled target DNA to a specific oligonucleotide on the membrane is taken to mean that the target has that specific sequence. This approach, and related DNA microarray methods, have many diagnostic applications.

Southern blot hybridization

Southern blotting was named after Edward M. Southern who developed this procedure in the 1970s. In this procedure, the target DNA is digested with one or more restriction endonucleases, size-fractionated by agarose gel electrophoresis, denatured and transferred to a nitrocellulose or nylon membrane for hybridization. During the electrophoresis, negatively charged DNA fragments, are sieved through the porous gel. Smaller DNA fragments move faster. For fragments between 0.1 and 20 kb long, the migration speed depends on fragment length, but scarcely at all on the base composition. Thus, fragments in this size range are fractionated by size in a conventional agarose gel electrophoresis system. To achieve efficient size-fractionation of large fragments (40 kb to several megabases), a more specialized system is required, such as a pulsed-field gel electrophoresis apparatus **(Fig 60)**.

Following electrophoresis, the test DNA fragments are denatured in strong alkali. As agarose gels are fragile, and the DNA in them can diffuse within the gel, it is usual to transfer the denatured DNA fragments by blotting on to a durable nitrocellulose or nylon membrane, to which single- stranded DNA binds readily.

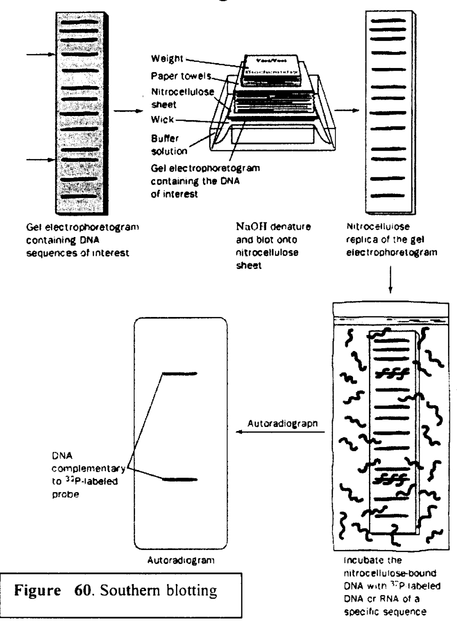


Figure 60. Southern blotting

The individual DNA fragments become immobilized on the membrane at positions which are a faithful record of the size separation achieved by agarose gel electrophoresis. Subsequently, the immobilized single-stranded target DNA sequences are allowed to associate with labeled single-stranded probe DNA. The probe will bind only to related DNA sequences in the target DNA, and their position on the membrane can be related back to the original gel in order to estimate their size.

Application: An important application of Southern blot hybridization in mammalian genetics is the ability to identify a given DNA probe as a member of a repetitive DNA family. Many important mammalian genes belong to multigene families, and many other DNA sequences show varying degrees of repetition. Once a newly isolated probe is demonstrated to be related to other uncharacterized sequences, attempts can then be made to isolate the other members of the family by screening genomic DNA libraries.

Southern blot hybridization has been used extensively in molecular genetic studies as a means of genomic restriction mapping: a labeled DNA probe from one genome can be used to infer the structure of related sequences in the same or different genomes. Because the genomic DNA samples are fractionated by separation of restriction fragments according to size, mutations which alter a restriction site, and significantly large insertions or deletions occurring between neighboring restriction sites, can be typed. Such mutations will result in altered restriction fragment lengths, that is **restriction fragment length polymorphisms (RFLPs)**.

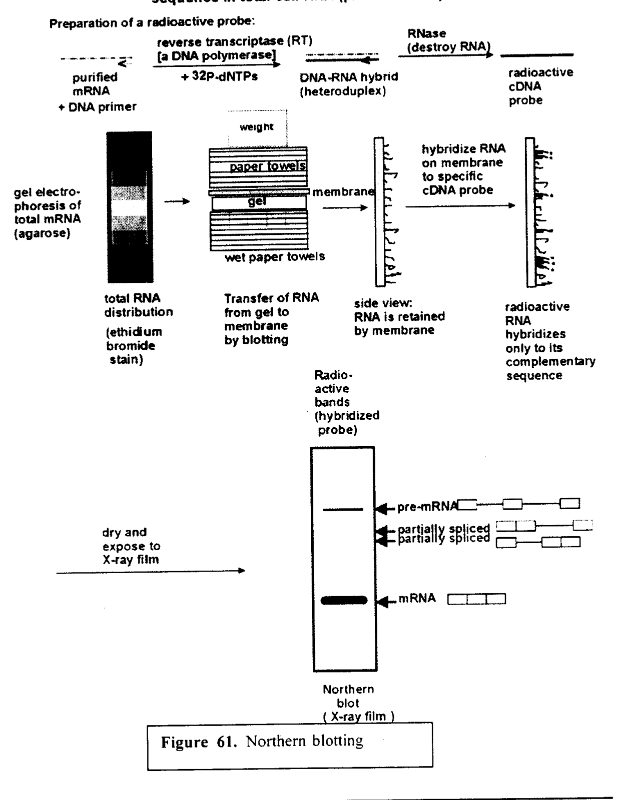


Figure 61. Northern blotting

Northern blot hybridization

Northern blot hybridization is a variant of Southern blotting in which the target nucleic acid is RNA instead of DNA. A principal use of this method is to obtain information on the expression patterns of specific genes. Once a gene has been cloned, it can be used as a probe and hybridized against a Northern blot containing, in different lanes, samples of RNA isolated from a variety of different tissues **(Fig. 61)**. The data obtained can provide information on the range of cell types in which the gene is expressed, and the relative abundance of transcripts. Additionally, by revealing transcripts of different sizes, it may provide evidence for the use of alternative promoters, splice sites or polyadenylation sites.

Detection of gene deletions by restriction mapping

Certain diseases are associated with a high frequency of deletion of all or part of a gene. If a partial restriction map has been established for the gene under investigation, deletions can be screened by Southern blot hybridization using an appropriate intragenic DNA probe. If the deletion is a small one, for example a few hundred base pairs, it is often apparent as a consistent reduction in size of normal restriction fragments in the gene. An individual who is homozygous for this mutation, or is a heterozygote with one normal allele and another with a small deletion, can easily be identified by detecting the aberrant size restriction fragments.

Large deletions will lead to absence of specific restriction fragments. Homozygous deletion of large DNA segments can easily be detected as complete absence of appropriate restriction

may still be detected by demonstrating comparatively reduced intensity of specific gene fragments. For example, patients with 21-hydroxylase deficiency often have deletions of about 30 kb of the 21-hydroxylase/C4 gene cluster. Such pathological deletions eliminate the functional 21-hydroxylase gene, CYP21, and an adjacent C4B gene, leaving the related CYP21P pseudogene and C4A genes. Patients with homozygous deletions will show absence of diagnostic restriction fragments associated with CYP21 and C4B, while carriers of the deletion will show a 2:1 ratio of CYP21P:CYP21 and of C4A:C4B.

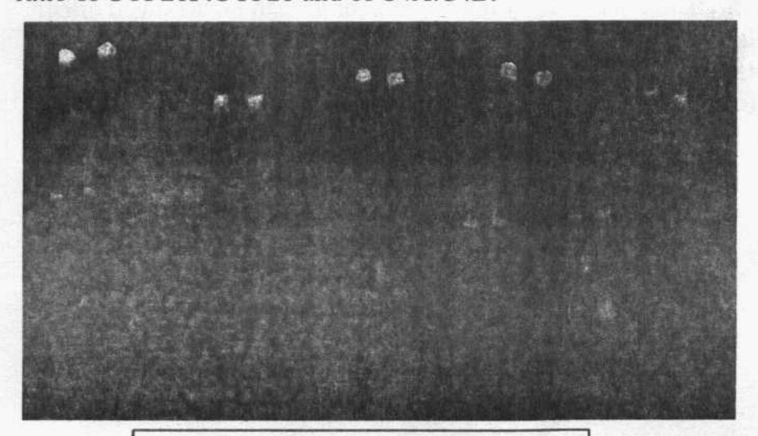
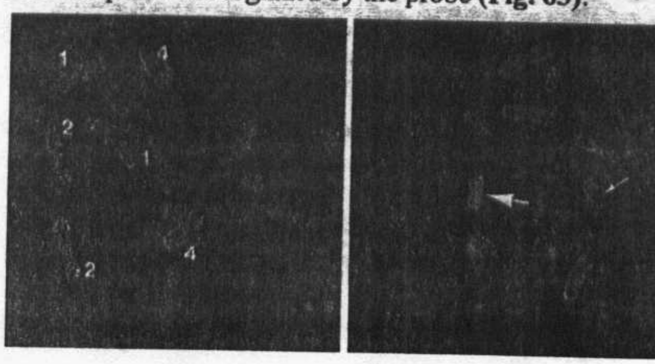


Figure 62. Chromosome in situ hybridization

Chromosome in situ hybridization

Chromosome in situ hybridization

A simple procedure for mapping genes and other DNA sequences is to hybridize a suitable labeled DNA probe against chromosomal DNA that has been denatured in situ. To do this, an air-dried microscope slide preparation of metaphase or prometaphase chromosomes is made, usually from peripheral blood lymphocytes **(Fig. 62)**. Treatment with RNase and proteinase K results in partially purified chromosomal DNA, which is denatured by exposure to formamide. The denatured DNA is then available for in situ hybridization with an added solution containing a labeled nucleic acid probe, overlaid with a coverslip. Depending on the particular technique that is used, chromosome banding of the chromosomes can be arranged either before or after the hybridization step. As a result, the signal obtained after removal of excess probe can be correlated with the chromosome band pattern in order to identify a map location for the DNA sequences recognized by the probe **(Fig. 63)**.



normal abnormal (translocation)

Figure 63. FISH technique

Tissue in situ hybridization

In this procedure, a labeled probe is hybridized against RNA in tissue sections. Tissue sections are made from either paraffin-embedded or frozen tissue using a cryostat, and then mounted on to glass slides. A hybridization mix including the probe is applied to the section on the slide and covered with a glass coverslip. Typically, the hybridization mix has formamide at a concentration of 50% in order to reduce the hybridization temperature and minimize evaporation problems **(Fig. 64)**..

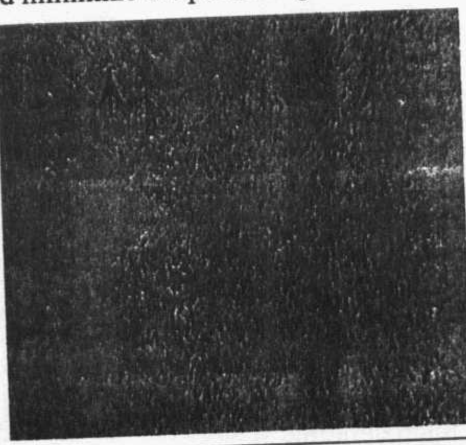


Figure 64. Sections of chicken oviduct and spleen hybridized with IgG γ-chain probes. Arrows indicate examples of IgG γ-chain mRNA-expressing cells. after hybridization with IgG γ-chain anti-sense probes.

DNA microarray technology

Recently developed **DNA microarrays (DNA chips) (Fig. 65)**. have provided a scale-up in hybridization assay technology because of their huge capacity for miniaturization and

automation. As in the case of reverse dot-blotting, the DNA microarray technologies employ a reverse nucleic acid hybridization approach: the probes consist of unlabeled DNA fixed to a solid support (the arrays of DNA or oligonucleotides) and the target is labeled and in solution. Although the technology for establishing DNA microarrays was only developed in the last few years, already there have been numerous important applications and their impact on future biomedical research and diagnostic approaches is expected to be profound.

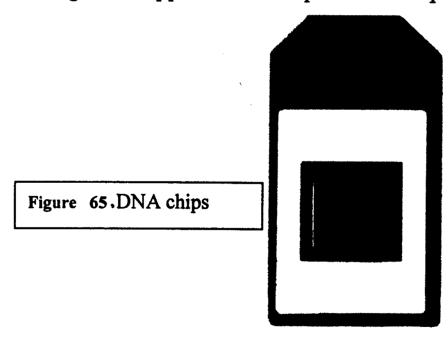


Figure 65. DNA chips

The microarrays can be divided into two main classes according to their method of construction:

1. Microarrays of DNA clones delivered by microspotting. Here the DNA clones have been prepared in advance and then printed onto the surface of a microscope slide.
2. Microarrays of oligonucleotides synthesized *in situ*. This approach has been pioneered by the

company Affymetrix, Inc., in Santa Clara, CA, and typically involves a combination of photolithography technology from the semiconductor industry with the chemistry of oligonucleotide synthesis .

Western blotting

Western blot analysis can detect **one** protein in a mixture of any number of proteins while giving information about the size of that protein. It does not matter whether the protein has been synthesized in vivo or in vitro. This method is, however, dependent on the use of a high-quality antibody directed against a desired protein. So you must be able to produce at least a small portion of the protein from a cloned DNA fragment. You will use this antibody as a probe to detect the protein of interest. Western blotting tells us how much protein has accumulated in cells.

Western blotting technique

1. **Separate the proteins** using SDS-polyacrylamide gel electrophoresis (also known as SDS-PAGE). This separates the proteins by size.
2. **Place a nitrocellulose membrane on the gel** and, using electrophoresis, drive the protein (polypeptide) bands onto the nitrocellulose membrane at a low pH.
3. **Incubate the nitrocellulose membrane with a primary antibody**. The primary antibody, which is the specific antibody mentioned above, sticks to your protein and forms an antibody-protein complex with the protein of interest.

4. Incubate the nitrocellulose membrane with a secondary antibody. This antibody should be an antibody-enzyme conjugate. The secondary antibody should be an antibody against the primary antibody. This means the secondary antibody will "stick" to the primary antibody, just like the primary antibody "stuck" to the protein. The conjugated enzyme is there to allow you to visualize all of this.

5. To actually see your enzyme in action, incubate it in a reaction mix that is specific for your enzyme. bands will appear wherever there is a protein-primary antibody-secondary antibody-enzyme complex, or, in other words, wherever the protein is.

6. Put x-ray film on your gel to detect a flash of light, which is given off by the enzyme. The reaction usually runs out in about an hour.

If you are interested in the rate of synthesis of a protein, Radio-Immune Precipitation (RIP) may be the best assay for you. Also, if a protein is degraded quickly, Western blotting would not detect it well; you will need to use (RIP). **Table 9** illustrates the differences between these two techniques.

Table 9 : differences between Western blotting and RIP

Western blotting	RIP
Not radioactive, so it's easier	Uses radioactivity
Tells how much protein has accumulated in a cell. Does not detect is rapidly degraded protein,	Tells the rate of synthesis of a protein.
Cannot detect protein-protein interactions.	Can detect protein-protein interactions. Two bands will show on a RIP gel. (Western blotting will only show one band.)

Summary

Nucleic acid hybridization is a fundamental tool in molecular genetics

Preparation of nucleic acid probes

DNA probes. Conventional DNA probes are isolated by

3- Cell-based DNA cloning.
4- PCR: PCR-derived DNA probes.

RNA probes can conveniently be generated from DNA which has been cloned in a specialized plasmid vector.

Oligonucleotide probes are short single-stranded pieces of DNA made by chemical synthesis.

Isotopic labeling and detection

Traditionally, by incorporating nucleotides containing radioisotopes (often ^{32}P, ^{33}P, ^{35}S or ^{3}H), which can be detected (autoradiography).

Nonisotopic labeling and detection

Fluorescence labeling is a popular nonisotopic labeling approach and detection is accomplished by fluorescence microscopy.

Dot-blot hybridization

simply spotting target DNA on a nitrocellulose or nylon membrane then allowing it to dry. exposed to a solution containing single-stranded labeled probe sequences. exposed to an autoradiographic film.

Allele-specific oligonucleotide (ASO) probes is a useful application of dot-blotting which distinguish between alleles that differ in a single nucleotide.

Reverse dot-blotting uses a reverse dot-blotting approach. This means that the target DNA is labeled and provided in solution.

Southern blot hybridization

the target DNA is digested with restriction endonucleases, size-fractionated by agarose gel electrophoresis, denatured and transferred to a nitrocellulose or nylon membrane for hybridization.

Northern blot hybridization

in which the target nucleic acid is RNA instead of DNA. A principal use of this method is to obtain information on the expression patterns of specific genes.

Chromosome in situ hybridization

hybridize a suitable labeled DNA probe against chromosomal DNA that has been denatured in situ. in order to identify a map location for the DNA sequences recognized by the probe.

Tissue in situ hybridization

In this procedure, a labeled probe is hybridized against RNA in tissue sections.

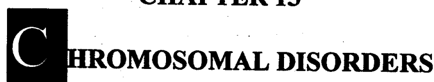
DNA microarray technology (DNA chips)

As in the case of reverse dot-blotting, the DNA microarray technologies employ a reverse nucleic acid hybridization approach: the probes consist of unlabeled DNA fixed to a solid support and the target DNA is labeled and in solution.

Western blotting

Western blot analysis can detect one protein in a mixture of any number of proteins while giving information about the size of that protein. This method depends on the use of a high-quality antibody directed against a desired protein.

CHAPTER 13



CHROMOSOMAL DISORDERS

Introduction

Chromosomes are the 46 rodlike structures seen during cell division in the nucleus of most human cells. They consist of proteins and DNA and carry most of the genes.

Preparing chromosomes for analysis is relatively simple. Circulating blood lymphocytes are generally used, except in the fetus, in whom amniocytes from amniotic fluid or chorionic villus cells from the placenta are used instead. The cells are cultured in the laboratory with phytohemagglutinin to stimulate cell division. Colchicine is added to arrest mitosis during metaphase, when each chromosome has replicated into two chromatids attached at the centromere. The cells, which are spread onto microscope slides, are then, stained. Chromosomes from single cells are usually photographed, their images cut out of the print and pasted onto a piece of paper, forming a karyotype. Computer imaging can also be used to produce a visual display of the chromosomes.

Chromosome staining is performed by G (Giemsa) or Q (fluorescent) banding techniques. Additional staining procedures and techniques for extending chromosome length have greatly increased the precision of cytogenetic diagnosis.

New molecular techniques use DNA probes (which can have fluorescent tags) to locate specific genes or DNA sequences

on the chromosomes. Fluorescent In Situ Hybridization (FISH) is used to identify the organization of genes and to look for deletions, rearrangements, and duplications of the chromosomes.

Karyotype nomenclature is as follows. The normal male is designated as 46,XY and the normal female as 46,XX. In Down syndrome due to an extra chromosome 21 (trisomy 21), the notation is 47,XY,21+ for a male and 47,XX,21+ for a female. In Down syndrome due to a translocation (two chromosomes that are stuck together), the typical 14/21 "balanced translocation carrier" mother is written as 45,XX, t(14q;21q). The translocation chromosome (t) is formed from 14q and 21q (in which q is the long arm); the short arms (p) are lost. For a deletion of the short arm of chromosome 5 (as in the 5p deletion syndrome), the female karyotype is 46,XX,5p.

Each arm of a chromosome is divided into one to four major regions, depending on chromosomal length; each band, positively or negatively stained, is given a number, which rises as the distance from the centromere increases. For example, 1q23 designates the chromosome (1), the long arm (q), the second region distal to the centromere (2), and the third band (3) in that region.

Autosomal abnormalities

Many different syndromes result from complete trisomies, partial trisomies (due to translocations of portions of either short or long arms), and deletions of the various chromosomes; only the more common clinical entities are included here.

Down Syndrome
(Trisomy 21; Trisomy G; Mongolism)

A chromosomal disorder usually resulting in mental retardation, a characteristic facies, and many other typical features, including microcephaly and short stature **(Fig.66)**.

The overall incidence among live births is about 1/800, but there is a marked variability depending on maternal age: for mothers < 20 yr, the incidence is about 1/2000; for mothers > age 40, it rises to about 1/40 overall. Just over 20% of infants with Down syndrome are born to mothers > 35 yr. Down syndrome may result from trisomy 21, translocation, or mosaicism.



Figure 66. Down syndrome

Trisomy 21: In about 95% of cases, there is an extra whole chromosome 21, which in 95% of these cases is maternally derived.

Translocation (t): Some persons with Down syndrome have 46 chromosomes but actually have the genetic material of 47 chromosomes; the additional chromosome 21 has been translocated or attached to another chromosome.

t(14;21) is the most common translocation, in which the additional chromosome 21 is attached to a chromosome 14. In about half the cases, both parents will have normal karyotypes,

indicating a de novo translocation in the affected child. In the other half, one parent (almost always the mother), although phenotypically normal, has only 45 chromosomes, one of which is t(14;21). Theoretically, the chance is 1:3 that a carrier mother will have a Down syndrome child, but for unknown reasons the actual risk is lower (about 1:10); if the father is the carrier, the risk is only 1:20.

t(21;22) is the next most common translocation. A carrier mother has about a 1:10 risk of having a Down syndrome child, with an even smaller risk for carrier fathers. In extremely rare cases, a parent carries a t(21;21). In this case, 100% of surviving offspring will be expected to have Down syndrome.

Mosaicism: Mosaicism occurs when two different cell types are present in a person. Down syndrome mosaicism presumably results from an error in chromosomal separation during cell division (nondisjunction) in the growing embryo. Most cases have two cell lines, one normal and one with 47 chromosomes. The relative proportion of each cell line is highly variable, both between persons and within different tissues and organs in the same person. The prognosis for intelligence presumably depends on the proportion of trisomy 21 cells in the brain. A few mosaic Down syndrome persons with barely recognizable clinical signs and normal intelligence have been identified. The incidence of mosaicism in Down syndrome is unknown. If a parent has germline mosaicism for trisomy 21, the risk of having a second affected child is increased.

Symptoms and Signs

Newborns tend to be placid, rarely cry, and demonstrate muscular hypotonia. Extra skin around the neck is common and

can be detected by fetal ultrasound as edema of the neck. Physical and mental development are retarded; the mean intelligence quotient (IQ) is about 50.

Microcephaly, a flattened occiput, and short stature are characteristic. The outer sides of the eyes are slanted upward, and epicanthal folds at the inner corner of the eye usually are present. Brushfield's spots (gray to white spots resembling grains of salt around the periphery of the iris) usually are visible and disappear during the first 12 months of life. The bridge of the nose is flattened, the mouth is often held open because of a large protruding tongue that is furrowed and lacks the central fissure, and the ears are small and rounded **(Fig. 67)**. Hands are short and broad and often have a single palmar crease (simian crease) **(Fig. 68)**; the fingers are short, with clinodactyly (incurving) of the 5th finger, which often has only two phalanges. The feet may have a wide gap between the 1st and 2nd toes, and a plantar furrow often extends backward on the foot. Hands and feet show characteristic dermal prints (dermatoglyphics).



Figure 67. Small and rounded ears



Figure 68. simian crease

Congenital heart disease, most commonly affecting the ventricular septum or the atrioventricular canal, occurs in about 40% of affected newborns. There is an increased incidence of

almost all other congenital anomalies, particularly duodenal atresia.

Many people with Down syndrome develop thyroid problems, which may be difficult to detect unless blood tests are done. Additionally, they are prone to developing hearing problems and problems with vision. Regular screening may be appropriate.

At autopsy, all adult Down syndrome brains show the typical microscopic findings of Alzheimer's disease, and many persons also develop the associated clinical signs. Some affected women are fertile; they have a 50% chance that their fetus will also have Down syndrome. However, many of these affected fetuses abort spontaneously. All men with Down syndrome are infertile.

Prognosis

Life expectancy is decreased because of heart disease and susceptibility to acute leukemia. Most affected persons survive to adulthood, but the aging process seems to be accelerated, with death often occurring in the 5th or 6th decade.

Trisomy 18
(Edwards Syndrome; Trisomy E)

A chromosomal disorder resulting from an extra chromosome 18, **(Fig. 69)**, producing many developmental abnormalities, including severe mental retardation.

Trisomy 18 occurs in 1/6000 live births, but many affected conceptions are miscarried. More than 95% of affected children have complete trisomy 18; 95% of extra chromosomes are

maternally derived. Advanced maternal age increases risk. Three females with trisomy 18 are born for every affected male.



Figure 69. Karyotype from a male with Edwards Syndrome (47,XY+18)

The affected newborn is markedly small for gestational age, with hypotonia and marked hypoplasia of skeletal muscle and subcutaneous fat. The cry is weak, and response to sound is decreased. There is often a history of feeble fetal activity, polyhydramnios, a small placenta, and a single umbilical artery. The orbital ridges are hypoplastic, the palpebral fissures short, the mouth and jaw small--all of which give the face a pinched appearance. Microcephaly, prominent occiput, low-set malformed ears, and a short sternum are common. A peculiar clenched fist with the index finger overlapping the 3rd and 4th fingers usually occurs. The distal crease on the 5th finger is absent, and there is a low-arch dermal ridge pattern on the fingertips. The fingernails are hypoplastic, and the big toe is

shortened and frequently dorsiflexed **(Fig. 70)**. Clubfeet and rocker-bottom feet are common. Severe congenital heart disease and anomalies of lungs, diaphragm, abdominal wall, kidneys, and ureters are common. Hernias and/or diastasis recti, cryptorchidism, and redundant skinfolds (particularly over the posterior aspect of the neck) also are common.





Figure 70. Edwards Syndrome

Survival for more than a few months is rare; < 10% are still alive at 1 yr. Marked developmental delay and disability are present in those who do survive.

Trisomy 13
(Patau Syndrome; Trisomy D)

A chromosomal disorder resulting from an extra chromosome 13, **(Fig. 71)**, producing many developmental abnormalities, including severe mental retardation and forebrain abnormalities.

Trisomy 13 occurs in about 1/10,000 live births; about 80% of cases are complete trisomy 13. Advanced maternal age increases risk, and the extra chromosome is usually maternally derived.

Figure 71. Karyotype from a female with Patau syndrome (47,XX+13)

Midline anomalies are characteristic. Gross anatomic defects of the brain, especially holoprosencephaly (failure of the forebrain to divide properly); cleft lip; cleft palate; microphthalmia; colobomas (fissures) of the iris; and retinal dysplasia are common. The supraorbital ridges are shallow, and the palpebral fissures usually are slanted. The ears are abnormally shaped and usually low-set. Deafness is common. Infants tend to be small for gestational age. Simian crease, polydactyly, and hyperconvex narrow fingernails are common. About 80% of cases have severe congenital cardiovascular anomalies; dextrocardia is common. Other midline defects include scalp defects and dermal sinuses. Loose folds of skin often present over the posterior aspect of the neck. The genitalia are frequently abnormal in both sexes; cryptorchidism and an abnormal scrotum occur in the male, and a bicornuate uterus occurs in the female. Apneic spells in early infancy are frequent. Mental retardation is severe **(Fig. 72)**.



Figure 72. Patau syndrome

Most patients (70%) are so severely affected that they die before age 6 mo; < 10% survive longer than 1 yr.

Deletion Syndromes

Clinical syndromes resulting from loss of parts of chromosomes.

5p-Deletion (cri du chat syndrome) involves a deletion of the end of the short arm of chromosome 5 and is characterized by a high-pitched, mewing cry, closely resembling the cry of a kitten, which is heard in the immediate newborn period, lasts several weeks, and then disappears. Affected newborns are of low birth weight and have microcephaly, facial symmetry and/or a round face with wide-set eyes, antimongoloid or downward-sloping palpebral fissures with or without epicanthal folds, strabismus, and a broad-based nose. The ears are low-set and abnormally shaped and frequently have narrow external auditory canals and preauricular tags. Varying degrees of syndactyly

occur; heart defects occur often and infants are hypotonic. Mental and physical development is markedly retarded. Many persons survive into adulthood but are very disabled **(Fig. 73)**.



Figure 73. Cri du chat syndrome

4p-Deletion (Wolf-Hirschhorn syndrome) results in profound mental retardation. There may be a broad or beaked nose, midline scalp defects, ptosis and colobomas, cleft palate, delayed bone age, and in males, hypospadias and cryptorchidism. A high mortality rate occurs during infancy; the relatively few survivors are prone to infections and epilepsy. The few who survive into their 20s are severely handicapped **(Fig.74)**.

Contiguous gene syndromes include microdeletions and submicroscopic deletions of the contiguous genes on particular parts of many chromosomes. They are detectable using fluorescent probes and techniques to extend chromosomes. Almost all cases are sporadic. Often the microdeletions cannot be

shown cytogenetically, but their presence can be confirmed by DNA probes specific to the deleted area.



Figure 74. Wolf-Hirschhorn syndrome

Telomeric deletions (deletions at either end of a chromosome) account for many nonspecific cases of mental retardation with dysmorphic features.

Sex chromosome abnormalities

Sex determination in humans is controlled by the X and the Y chromosomes. The female has two X chromosomes, and the male has one X plus one Y. The Y chromosome is among the smallest of the 46 chromosomes, and its major function seems to be

related to male sex determination. In contrast, the X is one of the largest chromosomes and contains hundreds of genes, most of which have nothing to do with sex determination.

Lyon hypothesis (X-inactivation): The normal female has two loci for every X-linked gene, as compared with the male's single locus. This would seem to produce a genetic "dosage" problem. However, according to the Lyon hypothesis, one of the two X chromosomes in each somatic cell of the female is genetically inactivated early in the life of the embryo. The Barr body, or sex chromatin mass within the nuclei of female somatic cells, represents the second inactivated X chromosome. The gene responsible for inactivating the genes of the X chromosome has been identified (XIST). Molecular genetic studies have demonstrated that not all genes on the second X chromosome are inactivated. In fact, no matter how many X chromosomes are present in the genome, all but one has most of the genes inactivated.

X-inactivation has interesting clinical implications. For example, X chromosome abnormalities are relatively benign, compared with analogous autosomal abnormalities. Females with three X chromosomes are often normal physically and mentally and are fertile. In contrast, all the known autosomal trisomies have devastating effects. Similarly, the absence of one X chromosome, although it leads to a specific syndrome (Turner syndrome), is relatively benign; the absence of an autosome is invariably lethal.

The asymptomatic carrier for X-linked recessive disorders may also be explained by X-inactivation. Females who are heterozygous for hemophilia or muscular dystrophy usually are

asymptomatic but occasionally show some bleeding tendencies or muscle weakness, respectively. The Lyon hypothesis suggests that X-inactivation is a random event; thus, in each person, 50% maternal and 50% paternal X-inactivation should occur. However, a random process follows the normal distribution curve, and most of the maternal X chromosomes may be inactivated in a specific tissue of some females or the paternal X chromosomes in other females. If, by chance, nearly all cells had the normal allele inactivated in a given tissue of a heterozygote, the disease in that person and in the hemizygous affected male would be similar.

Turner Syndrome
(Turner's Syndrome; Bonnevie-Ullrich Syndrome)

A sex chromosome abnormality in which there is complete or partial absence of one of the two sex chromosomes, producing a phenotypic female **(Fig.75)**.

Turner syndrome occurs in about 1/4000 live female births. Ninety-nine percent of 45,X conceptions are miscarried. Eighty percent of liveborn newborns with monosomy X have loss of the paternal X.

The chromosomal abnormalities in affected females vary. About 50% have a 45,X karyotype. Many patients are mosaics (eg, 45,X/46,XX or 45,X/47,XXX). The phenotype varies from that of a typical Turner syndrome to normal. Occasionally, affected persons have one normal X and one X that has formed a ring chromosome; for this to happen, a piece must be lost from both the short and the long arms of the abnormal X. Some affected persons have one normal X and one long arm isochromosome formed by the loss of short arms and

development of a chromosome consisting of two long arms of the X chromosome. These persons tend to have many of the phenotypic features of Turner syndrome; thus deletion of the short arm of the X chromosome appears to play an important role in producing the phenotype .

Affected newborns may present with marked dorsal lymphedema of the hands and feet and with lymphedema or loose folds of skin over the posterior aspect of the neck. However, many females with Turner syndrome are very mildly affected. Typically, short stature, webbing of the neck, low hairline on the back of the neck, ptosis, a broad chest with widely spaced nipples, multiple pigmented nevi, short 4th metacarpals and metatarsals, prominent finger pads with whorls in the dermatoglyphics on the ends of the fingers, hypoplasia of the nails, coarctation of the aorta, bicuspid aortic valve, and increased carrying angle at the elbow occur. Renal anomalies and hemangiomas are common. Occasionally, telangiectasia occurs in the GI tract, with resultant intestinal bleeding. Mental deficiency is rare, but many have some diminution of certain perceptual ability and thus score poorly on performance tests and in mathematics, even though they score average or above in verbal IQ tests. Gonadal dysgenesis with failure to go through puberty, develop breast tissue, or begin menses occurs in 90% of affected persons. Replacement with female hormones will bring on puberty. The ovaries are replaced by bilateral streaks of fibrous stroma and are usually devoid of developing ova. However, 5 to 10% of affected girls do go through menarche spontaneously, and very rarely, affected women have been fertile and have had children.



Figure 75. Turner syndrome

A cytogenetic analysis and Y-specific probe studies must be obtained for all persons with gonadal dysgenesis to rule out mosaicism with a Y-bearing cell line; eg, 45,X/46,XY. These persons are usually phenotypic females who have variable features of Turner syndrome. They are at high risk for gonadal malignancy, especially gonadoblastoma, and should have the gonads removed prophylactically as soon as the diagnosis is made.

The Triple X Syndrome (47,XXX)

A sex chromosome abnormality in which there are three X chromosomes, resulting in a phenotypic female.

About 1/1000 apparently normal females have 47,XXX. Other physical abnormalities are rare. Sterility and menstrual

irregularity sometimes occur. Mildly impaired intellect with IQ scores averaging just below 90 and associated school problems occur when compared with siblings. Advanced maternal age increases risk, and the extra X chromosome is usually maternally derived.

Rare Abnormalities of the X Chromosome

Although rare, 48,XXXX and 49,XXXXX females have been found. There is no consistent phenotype. The risk of mental retardation and congenital abnormalities increases markedly with an increase in the number of X chromosomes, especially when there are > 3. The genetic imbalance early in embryonic life before X-inactivation may cause anomalous development.

Klinefelter Syndrome
(Klinefelter's Syndrome; 47,XXY)

A sex chromosome abnormality in which there are two or more X chromosomes and one Y, resulting in a phenotypic male.

Klinefelter syndrome occurs in about 1/800 live male births. The extra X chromosome is maternally derived in 60% of cases.

Affected persons tend to be tall, with disproportionately long arms and legs. They often have small, firm testes, and about 1/3 develop gynecomastia. Puberty usually occurs at the normal age, but often facial hair growth is light. There is a predisposition for learning difficulties, and many have deficits in verbal IQ, auditory processing, and reading. Clinical variation is great, and many 47,XXY males are normal in appearance and intellect and are found in the course of an infertility workup (probably all 47,XXY males are sterile) or in cytogenetic surveys of normal

populations. Boys from the latter group have been followed developmentally. There is no increased incidence of homosexuality. Testicular development varies from hyalinized nonfunctional tubules to some production of spermatozoa, and urinary excretion of follicle-stimulating hormone is frequently increased **(Fig. 76)**.



Figure 76,Klinefelter syndrome

Variants: Mosaicism occurs in 15% of cases. Some affected persons have three, four, or even five X chromosomes along with the Y. In general, as the number of X chromosomes increases, the severity of mental retardation and of malformations also increases.

The 47,XYY Syndrome

A sex chromosome abnormality in which there are two Y chromosomes and one X, resulting in a phenotypic male.

The 47,XYY syndrome occurs in about 1/1000 males. Affected persons tend to be taller than average with a 10 to 15 point reduction in IQ for their family. There are few physical problems. Minor behavior disorders, hyperactivity, attention deficit disorder, and learning disabilities are increased.

Intersex states

Conditions in which the appearance of the external genitalia is either ambiguous or at variance with the person's chromosomal or gonadal sex.

Etiology and Classification

The genitalia form during the first 3 months of gestation via a cascade of events initiated by the fetal karyotype and mediated largely by the sex steroids. Aberrations in this cascade can produce genital ambiguities or inconsistencies, resulting in intersex states. Classification is most conveniently based on gonadal histology.

Female pseudohermaphrodites have ovaries and normal female internal genitalia but ambiguous external genitalia; they are genetically normal females with a 46,XX karyotype. The ambiguous external genitalia result from exposure to excessive amounts of androgens in utero. The offending androgen may be exogenous (eg, progesterone given to the mother to prevent miscarriage) but is more commonly endogenous, eg, resulting from an enzymatic block in steroidogenesis due to genetic aberrations on chromosome 6.

Male pseudohermaphrodites have gonadal tissue that is only testicular and usually have a 46,XY karyotype. The external genitalia are usually ambiguous, but this is variable and a female

phenotype is seen in the complete form of the testicular feminization syndrome (androgen insensitivity syndrome). Etiology is complex, but in general terms the disorder arises from inadequate production of androgen, inadequate response to androgen, or persistence of müllerian elements.

True hermaphrodites have both ovarian and testicular tissue, and mixed masculine and feminine genital structures depending on whether ovarian or testicular tissue predominates. In the USA, most true hermaphrodites have a 46,XX karyotype, but the pattern can be quite variable. Rarely, in true hermaphrodites the external genitalia is fully masculinized.

Patients with **mixed gonadal dysgenesis** have both testicular tissue and primitive gonadal tissue called streaks. These patients usually have a 46,XY/45,XO mosaic karyotype and ambiguous genitalia, and as adults they tend to be short of stature. When the streaks are bilateral, the disorder is termed **pure gonadal dysgenesis.** Such patients appear phenotypically as females.

Diagnosis

Patients with genital ambiguity, phenotypic females with palpable gonads, and phenotypic males with impalpable gonads should be evaluated for intersexuality. Males with lesser degrees of hypospadias usually do not require such an evaluation if both testes are descended and palpably normal.

Assessment of affected newborns is urgent not only to establish sex (particularly because of social pressures) but also to identify salt wasting, which may occur with adrenal virilism, before life-threatening hyponatremia develops. A blood sample

should be drawn immediately for karyotyping, but results may require several days.

Treatment

Assignment of appropriate sex is paramount. Generally, female pseudohermaphrodites are assigned as females. Male pseudohermaphrodites are assigned according to their genital development and hormonal activity. True hermaphrodites are also best assigned according to their genital development, but most have been reconstructed as males--an attractive option if the child has a normally descended testis to provide hormonal function at puberty. Male pseudohermaphrodites with full-blown testicular feminization syndrome must be assigned as females, but for many others, male assignment is appropriate. In marginal cases, 1 or 2 courses of testosterone propionate (in oil) 25 mg IM help determine the ability of the genitalia to respond to androgen--an essential requirement for male assignment.

Patients with mixed gonadal dysgenesis are best assigned as females, not only because of short stature but also because their gonads tend to develop tumors (gonadoblastoma). Early reconstruction of the external genitalia with gonadectomy is generally recommended. Patients with pure gonadal dysgenesis appear phenotypically as females and should be raised as such.

Optimal timing for reconstruction of the genitalia varies. Those assigned as females, other than those with adrenal virilism, should have resection of an enlarged clitoris as early as possible to facilitate acceptance by their family as females. Those with adrenal virilism should be deferred some months until they have been rendered endocrinologically stable by steroid therapy.

Vaginal reconstruction is best deferred until puberty because of the high incidence of stenosis when it is performed early in life. Correction of hypospadias in males usually is performed at about 1 to 2 yr of age.

Summary

Autosomal abnormalities
Many different syndromes result from complete trisomies, partial trisomies and deletions of the various chromosomes.

Down Syndrome (Trisomy 21; Trisomy G; Mongolism)
A chromosomal disorder usually resulting in mental retardation, a characteristic facies. The causes are **(1)Trisomy 21:** In about 95% of cases. (2)**Translocation (t):** Persons have 46 chromosomes but actually have the genetic material of 47 chromosomes; **t(14;21) t(21;22) and (3) Mosaicism:** Most cases have two cell lines, one normal and one with 47 chromosomes.

Trisomy 18(Edwards Syndrome; Trisomy E)
A chromosomal disorder resulting from an extra chromosome 18, producing

Trisomy 13(Patau Syndrome; Trisomy D)
A chromosomal disorder resulting from an extra chromosome 13, producing many developmental abnormalities, including severe mental retardation and forebrain abnormalities.

Deletion Syndromes

1. **5p-Deletion (cri du chat syndrome)** involves a deletion of the end of the short arm of chromosome 5 and is characterized by a high-pitched, mewing cry, closely resembling the cry of a kitten.
2. **4p-Deletion (Wolf-Hirschhorn syndrome)** results in profound mental retardation. There may be a broad or beaked nose, midline scalp defects, cleft palate, delayed bone
3. **Contiguous gene syndromes** include microdeletions and submicroscopic deletions of the contiguous genes on particular parts of many chromosomes. They are detectable using fluorescent probes and techniques to extend chromosomes. Often the microdeletions cannot be shown cytogenetically, but their presence can be confirmed by DNA probes specific to the deleted area.
4. **Telomeric deletions** (deletions at either end of a chromosome) account for many nonspecific cases of mental retardation with dysmorphic features.

Sex chromosome abnormalities

1. **Turner Syndrome(Bonnevie-Ullrich Syndrome):** A sex chromosome abnormality in which there is complete or partial

absence of one of the two sex chromosomes, producing a phenotypic female.

2. **The Triple X Syndrome (47,XXX):** A sex chromosome abnormality in which there are three X chromosomes, resulting in a phenotypic female. About 1/1000 apparently normal females have 47,XXX. Other physical
3. **Klinefelter Syndrome(47,XXY):** A sex chromosome abnormality in which there are two or more X chromosomes and one Y, resulting in a phenotypic male. **Variants:** Mosaicism occurs in 15% of cases. Some affected persons have three, four, or even five X chromosomes along with the Y. In general, as the number of X chromosomes increases, the severity of mental retardation and of malformations also increases.
4. **The 47,XYY Syndrome:** A sex chromosome abnormality in which there are two Y chromosomes and one X, resulting in a phenotypic male. There are few physical problems. Minor behavior disorders, hyperactivity, attention deficit disorder, and learning disabilities are increased.

Intersex states

1. **Female pseudohermaphrodites** have ovaries and normal female internal genitalia but ambiguous external genitalia. They are genetically normal females with a 46,XX karyotype. It is more commonly due to genetic aberrations on chromosome 6.
2. **Male pseudohermaphrodites** have gonadal tissue that is only testicular and usually have a 46,XY karyotype.
3. **True hermaphrodites** have both ovarian and testicular tissue, and mixed masculine and feminine genital structures depending on whether ovarian or testicular tissue predominates.
4. **Mixed gonadal dysgenesis** have both testicular tissue and primitive gonadal tissue called streaks. These patients usually have a 46,XY/45,XO mosaic karyotype

CHAPTER 14



GENE DISORDERS

Introduction

Human development depends on genetic and environmental factors. A person's genetic composition (genome) is established at conception. The genetic information is carried in the DNA of the chromosomes and mitochondria. Most diseases probably have some genetic component, the extent of which varies. Environmental factors may alter genetic information through mutation or other structural alteration and can affect classic genetic disorders (eg, dietary management in phenylketonuria, drugs for hypercholesterolemia).

Powerful molecular genetic techniques have made it possible to study the structure of DNA and to track changes during development and in different tissues. The structure of a gene is complex and includes control elements (eg, promoters, enhancers), expressed elements (exons), intervening elements that are not expressed (introns), and a termination signal. The configuration of the DNA of a gene in a tissue that does not express the gene is likely to be different (eg, methylated, condensed) than that of a gene that is actively expressing.

The human genome project is a large collaborative international project that began in 1991. The goal is to map specific genes to specific locations on chromosomes and determine their exact nucleotide sequence (the genome of a human) by 2005. Mapping can be done through family studies,

using the known location of DNA markers, and involves isolating small stretches of DNA and sequencing the genes and the other DNA in that area.

DNA's capacity to replicate constitutes the basis of hereditary transmission. DNA also provides the genetic code, which determines cell development and metabolism by controlling RNA synthesis. The sequence of the elements (nucleotides) that comprise DNA and RNA determines protein composition and thus its function.

Genes (between 60,000 and 100,000 in humans) are carried by the chromosomes (rodlike structures in the cell nuclei) and mitochondria (circular structures in cell cytoplasm present in multiple copies). In humans, somatic (nongerm) cells normally have 46 chromosomes, occurring as 23 pairs. Each pair consists of one chromosome from the mother and one from the father. One pair, the sex chromosomes, determines a person's sex. Women have two X chromosomes in every somatic cell nucleus, whereas men have one X and one Y chromosome (ie, heterologous chromosomes). The X chromosome carries genes responsible for many hereditary traits, whereas the small, differently shaped Y chromosome carries the genes that initiate male sex determination. The remaining 22 chromosome pairs, the autosomes, are usually homologous (ie, identical in size, shape, and position and number of genes). Germ cells (egg and sperm) undergo meiosis, which reduces the number of chromosomes to 23,half that of somatic cells (46), so that when an egg is fertilized by a sperm at conception, the normal number of chromosomes is reconstituted. In meiosis, the genetic information inherited from a person's mother and father is recombined through crossing over or exchange between the homologous chromosomes.

Genes, the basic units of heredity, are arranged linearly within the DNA along the chromosomes; each gene has a specific location (locus) or position on the chromosomes. The number and arrangement of loci on homologous chromosomes are usually identical. However, the structure of a specific gene may have minor variations (polymorphisms) without causing disease. The specific nucleotide sequence of the genes that occupy the two homologous loci along the two chromosomes of a pair are called alleles. The two alleles (ie, the one inherited from the mother and the one inherited from the father) may have slightly different nucleotide sequences or may be the same. A person with a pair of identical alleles for a particular gene is a homozygote; a person with a pair of dissimilar alleles is a heterozygote.

If a trait or disorder occurs when only one allele is abnormal, the disorder is said to be dominant. A disorder is said to be recessive if it occurs only when both alleles at the loci on both chromosomes are abnormal. A few genes are located in the mitochondrial DNA. Many copies of mitochondrial DNA are present in cytoplasm and may have the same DNA structure (homoplasmy) or different structures (heteroplasmy).

There are three types of genetic disorders: (1)Mendelian, or single-gene, mutations are inherited in recognizable patterns; (2) multifactorial conditions involve > 1 gene and environmental factors interacting in ways that are not always clearly recognizable but that have been described by observation (empirically); and (3) chromosomal abnormalities include structural defects and deviations from the normal number. More recently, mitochondrial and nontraditional patterns of inheritance have been recognized.

Humans vary phenotypically and genotypically. Heterogeneity results from different alleles and mutations of the multiple genes, which are part of almost all biochemical pathways. Mutations in different parts of a gene may cause different disorders. For instance, mutations in different locations along the gene for one type of collagen can produce tall stature, arthritis, and deafness or lethal dwarfism. More than 60 specific causes of congenital deafness have been identified; some are genetic (involving nuclear and mitochondrial genes), whereas others result from rubella virus or other environmental agents. Concern about the teratogenic effects of drugs taken during pregnancy is growing. For instance, women who consume alcohol during pregnancy are at increased risk of having infants with mental retardation and behavioral disturbances, intrauterine growth retardation, and congenital malformations.

Genetic Screening

Genetic screening may be used in populations at risk for a particular genetic disorder. Genetic screening is only appropriate when the natural history of the disease is understood; the screening tests are valid and reliable; sensitivity, specificity, false-negative, and false-positive rates are acceptable; and effective therapy is available. A sufficient benefit must be derived from a screening program to justify its cost.

Heterozygote screening: Screening a susceptible population (eg, Tay-Sachs disease in Ashkenazic Jews, sickle cell anemia in blacks, thalassemia in various ethnic groups) may be appropriate because of the high frequency of heterozygotes. Heterozygote screening can determine if a person is a carrier for a specific disorder. If the partner is also a heterozygote, the

couple is at risk of having an affected child. Screening allows the couple to make informed reproductive choices.

Presymptomatic genetic screening: Presymptomatic genetic screening may be appropriate for persons with a family history of a dominantly inherited disorder (eg, Huntington's disease, breast cancer). Identifying a definite carrier of the genetic disorder may allow the patient to make informed decisions (eg, monitoring in the case of breast cancer, reproductive choices in the case of Huntington's disease or adult polycystic kidney disease).

Prenatal diagnosis: Amniocentesis, chorionic villus sampling, umbilical cord blood sampling, maternal blood sampling, maternal serum screening, and fetal visualization with ultrasound and radiography are useful in prenatal diagnosis. Common reasons for prenatal screening include maternal age > 35 yr, a family history of a condition that can be diagnosed by prenatal techniques, abnormal maternal serum screening results, and certain complications of pregnancy.

Newborn screening: Screening for phenylketonuria, galactosemia, and hypothyroidism in the newborn allows prophylaxis (ie, special diet or replacement therapy) to be initiated early enough to prevent severe complications.

Construction of a Family Pedigree

The family history is often key to determining genetic risk. It is most easily recorded in a family pedigree (family tree), which uses conventional symbols. The pedigree provides a ready view of problems or illnesses within the family and facilitates analysis of inheritance patterns, including the range and degree of

affliction and variation among persons and generations. Some familial disorders with identical phenotypes have several patterns of inheritance. For example, cleft palate can have an autosomal dominant, autosomal recessive, or X-linked recessive pattern of inheritance, or it may be multifactorial (ie, familial but with no precisely predictable inheritance pattern).

Generations in the family pedigree are numbered with Roman numerals **(Fig. 77)** with older generations at the top and the most recent at the bottom. Within each generation, persons are numbered from left to right with Arabic numerals. Siblings are usually listed by age, with the oldest on the left. Thus, each member of the pedigree can be identified by two numbers (eg, II, 4). A spouse is also assigned an identifying number **(eg, II, 6 in Fig. 77).**

Study of a trait or disease begins with the affected person (the proband, propositus [male], proposita [female], or index case). When taking a family history, the physician usually draws the pedigree as the relatives are described. The inquiry begins with the siblings of the proband and proceeds to the parents; relatives of the parents, including brothers, sisters, nephews, and nieces; grandparents; and so on. The number of relatives included in the pedigree is determined by the inheritance pattern of the condition and by the extent of the informant's memory or knowledge. Usually, at least 3 generations are included. Illnesses, hospitalizations, causes of death, miscarriages, abortions, congenital anomalies, and any other unusual features are recorded.

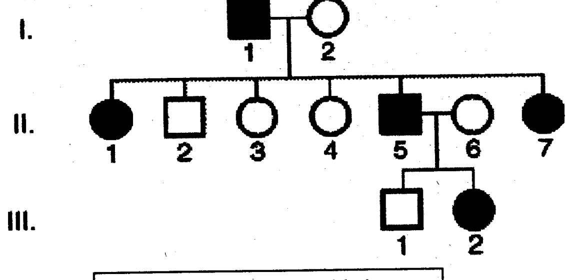


Figure 77 . Autosomal dominant inheritance.

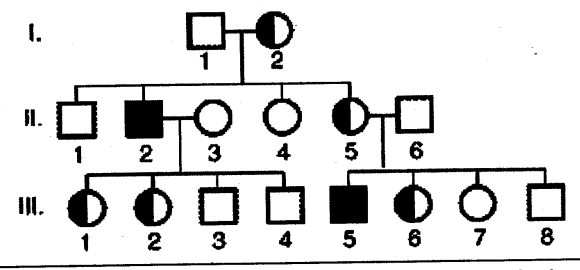


Figure 78. X-linked recessive inheritance. A shaded half-circle indicates a female carrier identifiable by testing.

Genetic Counseling

Genetic counseling involves obtaining a thorough family history and addressing the family's concerns and questions by gathering appropriate information from the literature and from genetic specialists; consultation with an expert concerning the specific

condition is often important. The information provided should include the diagnosis and diagnostic methods, including identification of carriers; the natural history of the disorder and its complications; the recurrence risk for the patient and various family members; potential therapies; and reproductive options. Communicating genetic risks and options is an involved process that often requires follow-up visits and written communication.

What are genetic disorders?

Both environmental and genetic factors have roles in the development of any disease. A genetic disorder is a disease caused by abnormalities in an individual's genetic material (genome). There are four different types of genetic disorders: (1) single-gene, (2) multifactorial, (3) chromosomal, and (4) mitochondrial.

(1) Single-gene (also called Mendelian or monogenic)

This type is caused by changes or mutations that occur in the DNA sequence of one gene. Genes code for proteins, the molecules that carry out most of the work, perform most life functions, and even make up the majority of cellular structures. When a gene is mutated so that its protein product can no longer carry out its normal function, a disorder can result. There are more than 6,000 known single-gene disorders, which occur in about 1 out of every 200 births. Some examples are cystic fibrosis, sickle cell anemia, Marfan syndrome, Huntington's disease, and hereditary hemochromatosis.

Single-gene disorders are inherited in recognizable patterns: autosomal dominant, autosomal recessive, and X-

linked. More information on the different modes of inheritance is available from the following Web sites:

Inheritance Of Single-Gene Defects

Genetic disorders determined by a single gene are easiest to analyze and the most fully studied. Many specific disorders have been described. Single-gene defects may be autosomal or X-linked, dominant or recessive.

Autosomal dominant

A person need have only one abnormal allele of a gene for it to lead to an autosomal dominant disorder. A typical pedigree of an autosomal dominant trait is shown in **Fig.77** vertical transmission occurs. In general, the following rules apply:

- An affected person has an affected parent.
- An affected person and an unaffected person have, on average, an equal number of affected and unaffected children.
- Unaffected children of an affected parent have unaffected children and grandchildren.
- Males and females are equally likely to be affected.
- The risk for occurrence among children of an affected person is 50%.

Pedigrees are based on the phenotype (observable features). With molecular studies, the genotype can also be determined and recorded.

Expressivity and penetrance: A gene's effects may be influenced by the environment and by other genes that may alter

phenotypic expression (or expressivity). Thus, even within a family, persons with the same gene alteration (allele) may manifest widely variable phenotypes. For instance, in Waardenburg syndrome type I, a mutation in the *PAX3* gene typically results in a white forelock, wide-spaced eyes, and heterochromia of the iris, but < 20% of cases have significant hearing loss. Some family members with the abnormal allele have none of the features except a white forelock, but their children have serious congenital deafness due to the same mutation. Rarely, expressivity is so low that a clinical abnormality cannot be detected, yet the unaffected carrier of the abnormal allele can pass it to offspring, who may develop the full clinical picture. In this case, the pedigree appears to skip a generation. This phenomenon is known as lack of penetrance. However, some cases of apparent lack of penetrance are due to the examiner's failure to recognize or be familiar with minor manifestations of the condition. Cases with minimal expressivity are sometimes referred to as a forme fruste of the disease.

Pleiotropy: A single-gene defect may produce multiple anomalies in different organ systems. For example, breakable bones, deafness, bluish whites of the eyes, dysplastic teeth, hypermobile joints, and heart valve abnormalities may occur in osteogenesis imperfecta (an abnormality of connective tissue in which many patients have recognizable abnormalities of collagen genes. Because all of these clinical anomalies involve collagen in various tissues and because specific collagen types have a predictable distribution, many organs and systems are involved.

Sex-limited inheritance: A trait that appears in only one sex is called sex-limited. This is different from X-linked inheritance, which refers to traits carried on the X chromosome.

Sex hormones and other physiologic differences between males and females may alter the expressivity of a gene. For example, premature baldness is an autosomal dominant trait, but presumably as a result of female sex hormones, the condition is rarely expressed in the female, and then usually only after menopause. Thus, sex-limited inheritance, perhaps more correctly called sex-influenced inheritance, is a special case of limited expressivity and penetrance.

Mutations: Mutations are changes in genetic information that arise spontaneously. An autosomal dominant pedigree begins with a fresh mutation, in which the genetic information (DNA) inherited from the parents has been changed. Rates of mutation vary (1/6,000 to 1/50,000 persons) among recognized autosomal dominant disorders.

For example, about 80% of persons with achondroplastic dwarfism (an autosomal dominant trait) have no family history and thus represent new mutations. Almost all affected persons can transmit the newly mutated gene to their offspring. In achondroplasia, mutations occur at a very specific place in the gene. In many other conditions, each new mutation seems to occur at a different place in the gene.

In general, unaffected parents are not at increased risk of having additional affected offspring, but a few normal-appearing parents have had two or even three offspring with typical achondroplastic dwarfism. The explanation is a germline mutation; ie, an event early in the embryonic life of a normal-appearing parent when only a handful of germ cell precursors are present. The cell possessing the new mutation can then contribute many cells to the developing gonad. In such cases, the chance of

having another affected child may be as high as 50%. The existence of germline mutations has been confirmed through molecular studies of the gene mutation in parents and offspring. In the case of the phenotypically normal father who carries a germline mutation, molecular studies have confirmed that many sperm may carry the mutation.

Autosomal recessive

A typical pedigree is shown in **Fig 79**. A person must have two copies of an abnormal allele to develop an autosomal recessive disorder. Certain populations may have an increased rate of heterozygotes or carriers because of a founder effect (eg, the group started with few members, one of whom was a carrier) or because ancestry may have given a carrier a selective advantage (eg, sickle cell anemia, in which being a heterozygote protects against malaria). In general, the following rules of inheritance apply:

- Normal parents have an affected child, both parents are heterozygotes, and, on average, 1/4 of their children will be affected, 1/2 will be heterozygotes, and 1/4 will be normal.
- All children of an affected person and a genotypically normal person will be phenotypically normal heterozygotes.
- On average, 1/2 the children of an affected person and a heterozygote will be affected, and 1/2 will be heterozygotes.

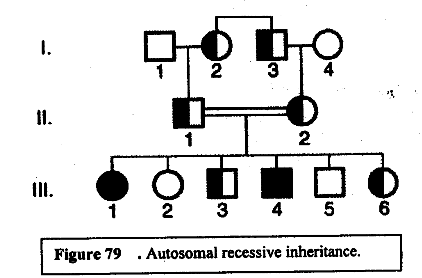


Figure 79 . Autosomal recessive inheritance.

- All children of two affected persons will be affected.
- Males and females are equally likely to be affected.
- Heterozygotes are phenotypically normal but are carriers of the trait. If the disease is caused by a defect of a specific protein (eg, an enzyme), the carrier usually has a reduced amount of that protein. If the mutation is known, molecular genetic techniques can identify heterozygous phenotypically normal persons.

Consanguinity (eg, mating of related persons) may be important in autosomal recessive diseases. Related persons are more likely to share the same mutant allele. It is estimated that each human is a heterozygote (ie, a carrier) for six to eight alleles that would, in the homozygous state, lead to disease. A detailed family history may disclose unknown or forgotten consanguinity. Parent-child or brother-sister unions (usually referred to as incest) have an increased risk for abnormal offspring, because 50% of their genes are the same.

X-linked dominant

A typical pedigree is shown in **Fig 80** . In general, the following rules of inheritance apply:

- Affected males transmit the trait to all of their daughters but not their sons (male-to-male transmission does not occur).
- Affected heterozygous females transmit the condition to 1/2 of their children, regardless of sex.
- Affected homozygous females transmit the trait to all of their children.
- Twice as many affected females as males will have the disorder unless it is lethal in males.

X-linked dominant conditions, which are very rare, usually affect males more severely than females; however, females carrying only one abnormal allele are affected. For example, in incontinentia pigmenti, the X-linked gene appears to be lethal in males and causes a peculiar swirling pattern of melanin pigmentation and other patchy anomalies of teeth, eyes, and CNS in females. In nephrogenic diabetes insipidus, females show only mild degrees of polydipsia and polyuria.

X-linked dominant inheritance is difficult to differentiate from autosomal dominant inheritance without the use of molecular probes. Large pedigrees are required, with particular attention given to the offspring of affected males, because male-to-male transmission rules out X-linkage (males pass their Y chromosomes to their sons).

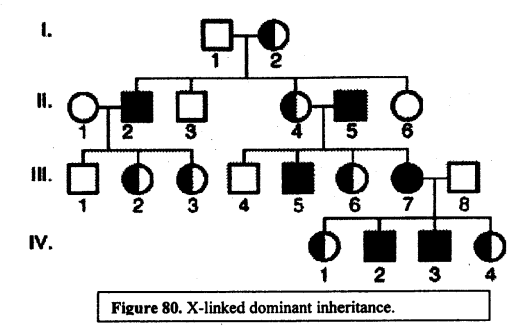


Figure 80. X-linked dominant inheritance.

X-linked recessive

A typical pedigree is shown in **Fig 81**. . In general, the following rules of inheritance apply:

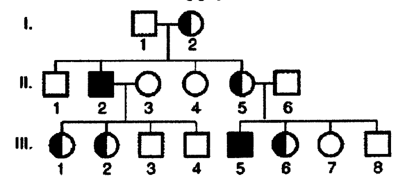


Figure 81. X-linked recessive inheritance. A shaded half-circle indicates a female carrier identifiable by testing.

- Nearly all affected persons are male.
- If the trait is transmitted through the heterozygous mother, she is usually phenotypically normal

- The trait may represent a new mutation in the affected male.
- An affected male never transmits the trait to his sons.
- All daughters of an affected male will be carriers.
- The carrier female transmits the trait to 1/2 of her sons.
- No daughters of a carrier female will show the trait, but 1/2 will be carriers (unless they also inherit the trait from their father, such as color blindness).

A female must have the abnormal gene on both X chromosomes (homozygote) for it to be expressed. This can happen if her father is affected and her mother is heterozygous or homozygous for the mutant allele. However, in males, all genes on the X chromosome, whether recessive or dominant, are expressed. Because most X-linked recessive disorders are rare, affected females are very rare (the incidence in females is the square of that in males). In addition, on average, half of the maternal uncles of the proband are affected; also, half of the maternal aunts are carriers, so some of the proband's male maternal first cousins will be affected.

Occasionally, females who are heterozygous for X-linked mutations show varying degrees of expression, but they are rarely affected as severely as hemizygous males. A structural chromosomal rearrangement (eg, an X-autosome translocation, a missing or deleted X chromosome) can allow a female to be affected even though she is a heterozygote.

Codominant

In codominant inheritance, both alleles are expressed in a heterozygous person. Because usually both alleles at a genetic locus produce some product, codominance, or the expression of both alleles, is noticed only if the phenotypes or products of the alleles are qualitatively different, as with the blood group antigens (eg, AB, MN), the WBC antigens (eg, DR4, DR3), and serum proteins differing in electrophoretic mobility (eg, albumin, haptoglobulin).

(2) Multifactorial (also called complex or polygenic)

This type is caused by a combination of environmental factors and mutations in multiple genes. For example, different genes that influence breast cancer susceptibility have been found on chromosomes 6, 11, 13, 14, 15, 17, and 22. Its more complicated nature makes it much more difficult to analyze than single-gene or chromosomal disorders. Some of the most common chronic disorders are multifactorial disorders. Examples include heart disease, high blood pressure, Alzheimer's disease, arthritis, diabetes, cancer, and obesity. Multifactorial inheritance also is associated with heritable traits such as fingerprint patterns, height, eye color, and skin color.

Multifactorial Inheritance

Close relatives (up to third-degree) tend to resemble each other with respect to a number of quantitative or measurable characteristics (eg, height, weight, size and shape of nose, facial features, BP, intelligence). Many traits are distributed along a bell-shaped curve, a phenomenon that is compatible with determination of the trait by several genes. Each gene adds to or

subtracts from the trait and acts in an additive manner independent of other genes. Few persons are at the extremes of the distribution and many are in the middle, because a person is unlikely to inherit many factors acting in the same direction. Environmental factors, each adding to or subtracting from the final result, will also produce a normal distribution.

Many relatively common congenital anomalies and familial diseases do not follow expectations for single-gene (Mendelian) inheritance. More likely, these conditions result from multifactorial inheritance; a threshold separates the affected person from the unaffected person. The affected person is predisposed to the condition, representing the sum of genetic and environmental influences. Thus, the risk of expression of such a trait in first-degree relatives (siblings and children), who share 50% of the affected person's genes, is relatively high. The risk in more distant relatives, who are likely to inherit only a few high-liability genes, is much lower.

For example, Neural Tube Defects (NTDs, anencephaly, spina bifida, encephalocele, myelomeningocele) usually have multifactorial causes, although they also have many specific known single-gene, chromosomal, and environmental causes. In the North American white population, NTDs occur with a combined incidence of about 1.5/1000 live births. Unaffected parents of an affected infant are thought to carry many high-liability genes and have about a 1/30 (3%) chance of having a second affected offspring. Similarly, a parent with a multifactorial NTD has a 3 to 4% chance of having an affected child. In the rare instance when a couple has had two affected children, the risk for a third increases to 7 to 8%.

Environmental factors play a role in multifactorial inheritance. The incidence of NTDs approaches 1/100 in western parts of the United Kingdom, yet when persons from a high-risk area emigrate to North America, their risk falls but remains higher than that of the North American population. Also, 50 to 70% of NTDs can be prevented by maternal folic acid supplementation (400 μg/day) 1 mo. before conception to 3 mo. after conception. Although maternal folic acid deficiency plays a major role, it is not the only environmental cause of NTDs.

Other examples of multifactorial inheritance, with similar risks for siblings and offspring of affected persons, are congenital anomalies of the heart, idiopathic epilepsy (petit or grand mal), and most cases of cleft lip with or without cleft palate. In congenital pyloric stenosis, the male:female ratio is 5:1, suggesting that the threshold for females is higher. Thus, compared with a male, a female requires more potent liability genes to develop the condition, has more affected siblings, and has a greater risk of having affected offspring.

Attention is being focused on common adult disorders with multifactorial causes (eg, hypertension, arteriosclerotic heart disease, diabetes mellitus, cancer, arthritis). Many specific genes are being found. Genetically determined predisposing factors, including a family history and biochemical and molecular parameters, can identify persons at risk who would most likely benefit from preventive measures.

Nontraditional Inheritance

In addition to the traditional inheritance patterns described above, several other mechanisms have been described. They involve

genetically determined structures and can produce various disorders.

Mosaicism: Mosaicism is the presence of 2 cell lines differing in genotype or karyotype but derived from one zygote. Mutations are likely to occur during the cell divisions of any large multicellular organism even though the genetic apparatus of cell division is usually accurate and many mechanisms exist to repair mistakes made during replication. It is estimated that each time a cell divides, four or five changes occur in the genome. Thus, any large multicellular organism will have subclones of cells having a slightly different genetic makeup. These somatic mutations (ie, mutations during mitotic cell division) may not lead to disease but to disorders in which patchy changes occur. Molecular genetic techniques have shown mutations in the abnormal cells involved in a patch compared with the normal surrounding supporting tissues. For example, in McCune-Albright syndrome, there are patchy dysplastic changes in the bone, abnormalities of endocrine glands, patchy pigmentary changes, and occasionally abnormalities in the heart or liver. Persons with these abnormalities in all cells would die, so the condition would not be passed on to the next generation, but they survive because normal tissue supports the abnormal tissue. Occasionally, in a single-gene disorder, a parent seems to have a milder form but is actually a mosaic; their more severely affected child would have received a germ cell with the mutant allele and thus would have the abnormality present in every cell. Chromosomal mosaicism occurs in some embryos and can be demonstrated in the placenta on chorionic villus sampling. Most chromosomally abnormal embryos and fetuses abort spontaneously. However, the development of normal cells may

support certain chromosomal abnormalities, allowing offspring to be born alive.

Genomic imprinting: Genomic imprinting is the differential expression of genetic material depending on whether it has been inherited from the father or mother. Genomic imprinting is tissue-specific and time-in-development-specific. Bi-allelic or biparental expression of alleles may be present in some tissues and uniparental expression in other tissues. Angelman syndrome and Prader-Willi syndrome can both be produced by deletions of chromosome 15. Groups of specific genes exist in close proximity on chromosome 15 with only paternal or maternal expression. Depending on whether the deleted chromosome is paternally or maternally inherited, a different syndrome will be produced.

Many areas on several chromosomes have this type of parent-of-origin effect. The involved genes seem to be related to growth and behavior in early development. Some of these genes are also involved in tumors and cancers. Genomic imprinting must be considered in disorders that appear to have skipped a generation.

Uniparental disomy: Uniparental disomy occurs when two chromosomes of a pair are inherited from only one parent. This is very rare and is thought to involve trisomy rescue; ie, the zygote started off as a trisomy and one of the three chromosomes was lost, leading to uniparental disomy in 1/3 of cases. Imprinting effects may be seen because genetic information from the other parent is absent. In addition, if the same chromosome is in duplicate (isodisomy) and that chromosome carries an abnormal allele for an autosomal recessive disorder, a person

with uniparental disomy can have an autosomal recessive disorder although only one parent is a carrier. Vestigial chromosomal abnormalities in some tissues must be considered in the presence of uniparental disomy.

Triplet repeat, unstable mutations: A triplet repeat is an unusual type of mutation in which a triplet of nucleotides increases in number within a gene (a normal gene has relatively few tandem triplet repeats). This type of mutation has been recognized to occur in several disorders, particularly those involving the CNS. When the gene is transmitted from one generation to the next, or sometimes within the body as cells divide, the triplet repeat can expand and enlarge to a point at which the gene stops functioning normally. Examples include myotonic dystrophy, Huntington's disease, fragile X mental retardation, and several other neurologic disorders. The number of repeats may increase dramatically in the formation of germ cells or in certain tissues as the embryo and fetus develop. Expansion may be greater when transmitted from one parent (eg, the mother in myotonic dystrophy, the father in Huntington's disease); thus, a parent-of-origin effect and anticipation can be observed. This type of mutation is detected by molecular studies.

Anticipation: Anticipation occurs when a disorder has an earlier age of onset and severity of expression in each successive generation. It may occur because a parent is a mosaic, and the child has the full mutation in all cells. Triplet repeat expansion may demonstrate anticipation when the number of repeats increases with each generation.

(3) Chromosomal

Chromosomes, distinct structures made up of DNA and protein, are located in the nucleus of each cell. Because chromosomes are carriers of genetic material, such abnormalities in chromosome structure as missing or extra copies or gross breaks and rejoinings (translocations), can result in disease. Some types of major chromosomal abnormalities can be detected by microscopic examination. Down syndrome or trisomy 21 is a common disorder that occurs when a person has three copies of chromosome 21.

Congenital anomalies

Congenital anomalies may be inherited or sporadic, isolated or multiple, apparent or hidden, gross or microscopic. They cause nearly half of all deaths in term newborns. A major anomaly is apparent at birth in 3 to 4% of newborns; up to 7.5% of children manifest a congenital defect by age 5 yr.

Incidence varies with the type of defect; the geographic area, presumably due to genetic and/or environmental factors (spina bifida occurs in 3 to 4/1000 births in areas of Ireland but in 1/1000 in the USA); and cultural practices (consanguineous marriages increase the risk of genetic abnormalities). Increasing age of the mother (and, to a lesser extent, of the father) may increase the risk of chromosomal defects, especially Down syndrome.

Etiology may involve genetic and/or teratogenic factors. Different factors operating at the same period of organogenesis may produce identical defects. Genetic factors may cause many single anomalies and syndromes. They may operate via simple

mendelian or multifactorial inheritance. Some syndromes, such as Down syndrome, result from chromosomal abnormalities. Teratogenic factors include environmental toxins, radiation, diet, drugs, infection, and metabolic disorders.

Diagnosis

Prenatal diagnosis may be possible by ultrasonography, amniocentesis, or chorionic villus sampling.

Treatment

If a defect is identified prenatally and is serious, parents can decide whether to terminate the pregnancy.When an anomaly is identified at or after birth, parents should be informed promptly, although extensive discussion may be deferred until specialists are consulted. The family should be given a realistic appraisal of the severity of the condition, its prognosis, and the medical care available, and they should actively participate in decisions. If genetic factors are suspected, the parents should receive genetic counseling.

Congenital Heart Disease Anatomic defects of the heart and great vessels produced at various stages of fetal development and present at birth.

Cleft lip and **cleft palate** are the most common 1st arch defects, occurring once in 700 to 800 births. Among the postulated causes are the use of benzodiazepines during early pregnancy. The cleft may vary from involvement of the soft palate only, to a complete cleft of the soft and hard palates, the alveolar process of the maxilla, and the lip. The mildest form is a **bifid uvula.** An isolated cleft lip is primarily a cosmetic problem.

A cleft palate interferes with feeding and speech development. Early treatment consists of special cleft palate nipples or dental appliances that occlude the cleft so suckling can occur, and use of a feeder with which formula can be delivered with mild pressure (eg, a plastic bottle). Ultimate treatment is surgical closure; however, surgery may interfere with growth centers surrounding the premaxilla. Plastic surgery can significantly improve this disorder, but with inadequate treatment there is a nasal voice, compromised appearance due to mid-face deficiency, and a tendency to regurgitate. Dental, orthodontic, psychiatric, and speech therapy may be required.

Congenital torticollis is head tilt present at birth. The most common cause is traumatic neck injury during (but sometimes before) delivery, with hematoma, fibrosis, and contracture of the sternocleidomastoid (SCM) muscle. Other causes include abnormalities of the bony spine, such as Klippel-Feil syndrome (fusion of the cervical vertebrae) or of the atlas to the occipital bone (atlanto-occipital fusion). CNS tumors, bulbar palsies, and ocular dysfunction are prominent neurologic causes but are rarely present at birth. (See also Spasmodic Torticollis in Ch. 59.) Fractures, dislocations, or subluxations of the cervical spine (especially C-1 and C-2) or odontoid abnormalities are rare but serious causes, since permanent neurologic damage may result. Torticollis due to birth trauma is not present at birth; it appears in the first few days or weeks of life, and a nontender mass is noted in the SCM, usually in the midsegment. Cervical imaging studies are important to exclude bony causes, which may require stabilization. When torticollis is due to birth trauma, passive SCM stretching (by rotating the head and flexing the neck laterally to the opposite side) is indicated.

Hip, leg, and foot abnormalities

Congenital dislocation of the hip is more common in female than male infants and in infants with breech presentation. It seems to be secondary to laxity of the ligaments around the joint or to in utero positioning. The dislocation can be uni- or bilateral. If unilateral, the involved leg is shorter and there may be asymmetric skin creases in the thigh. The major sign of subluxation or dislocation is inability to completely abduct the thigh when the hip and knee are flexed. This is due to adductor spasm, which is often present even if the hip is not actually dislocated at the time of examination. If the hip is dislocated, abduction and external rotation of the femur may produce an audible or palpable "clunk" as the femoral head reenters the acetabulum (Ortolani's sign). Minor "clicks" are more common. They may resolve within a month or two but should be followed closely. Partial or complete dislocation may be difficult to detect at birth; periodic testing for limitation of hip abduction during the first year of life is advised. Hip ultrasonography appears accurate in establishing early diagnosis. However, hip x-rays may be difficult to interpret early and are helpful only if they confirm the clinical impression. Early treatment is critical, since the hip usually can be reduced immediately after birth, and with growth the acetabulum will then form almost normally. However, if therapy is delayed, the potential for correction without surgery declines steadily. Treatment consists of devices (eg, splints, slings, harnesses, or large padded diapers) that hold the affected hips abducted and externally rotated, thus encouraging the acetabulum to form properly with growth.

Femoral torsion or twisting, either internal (anteversion--knees pointing toward each other) or external (retroversion--

knees pointing in opposite directions), is typical in newborns, in whom either condition may be striking. Spontaneous correction of even dramatic femoral torsion generally occurs when the infant stands and walks. Sleeping prone can prolong retroversion. Hip x-rays or ultrasonography for dislocation should be considered. Positioning and passive exercises may be helpful.

Knee dislocation anteriorly with hyperextension at birth is rare but requires emergency treatment. The dislocation may be related to muscle imbalance (if myelodysplasia or arthrogryposis is present) or intrauterine positioning. It is often associated with ipsilateral hip dislocation. If the infant is otherwise normal, immediate treatment with daily passive flexion movements and splinting in flexion usually results in a functional knee.

Bowing and twisting (torsion) of the tibia are common at birth and are seldom pathologic. Bowing with x-ray changes of a narrow sclerotic intramedullary canal (Blount disease) is an exception, with a high risk of fracture and pseudarthrosis; a protective orthosis is needed.

In **talipes equinovarus,** the most common of the clubfoot (talipes) deformities, the foot is plantar flexed, inverted, and markedly adducted. Deformities from in utero positioning can mimic clubfoot, but they can be passively corrected, whereas pathologic changes cannot. Orthopedic care, beginning in the nursery with repeated cast applications to normalize the foot's position, is optimal. In severe cases, surgery may be required if casting is not successful.

In **talipes calcaneovalgus,** the foot is flat or convex, is dorsiflexed, and can easily be approximated against the lower tibia. Early treatment with a cast to place the foot in the

equinovarus position or with corrective shoes usually is successful.

Abnormalities of bone and cartilage

The **chondrodystrophies** are diseases that affect the way in which cartilage is converted to bone. Of these, achondroplasia is the best known. All are characterized by dwarfism (usually with a normal-sized trunk but with short extremities) and are often associated with abnormalities elsewhere. Mental development usually is normal. In most chondrodystrophies, x-rays of the long bones are needed for accurate diagnosis. Hypothyroidism should be ruled out. Treatment is supportive.

The **mucopolysaccharidoses** are similar to the chondrodystrophies, but some types also have visceral and CNS involvement with mental deficiency. Truncal shortening and limb contractures are often present. Bony changes are not apparent in newborns and develop as growth occurs.

Osteogenesis imperfecta, or "brittle-bone" disease, is a defect in the production of collagen resulting in repeated fractures with minor trauma. The severity varies widely, with the neonatal (congenital) type being the most severe. Fractures can occur in utero and during delivery. Trauma during delivery may lead to intracranial hemorrhage and stillbirth. At birth the skull is soft and feels like a "bag of bones." Infants born alive may die suddenly during the first few days or weeks. Survivors develop shortened extremities and other bony deformities. Mental development is normal unless CNS injury occurs. The sclerae are thin, translucent, and appear blue. Hearing loss from otosclerosis may occur. Orthopedic treatment, physiotherapy, and occupational therapy are directed at preventing fractures and

increasing function. Treatment with the bisphosphonate pamidronate given IV has been shown to increase bone mineral density and motor function and may decrease bone pain and bone resorption (**Fig. 82**).

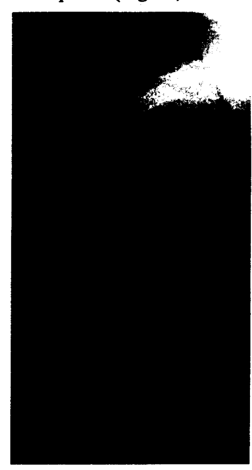


Figure 82. Osteogenesis imperfecta

(4) Mitochondrial

This relatively rare type of genetic disorder is caused by mutations in the nonchromosomal DNA of mitochondria. Mitochondria are small round or rod-like organellės that are involved in cellular respiration and found in the cytoplasm of

plant and animal cells. Each mitochondrion may contain 5 to 10 circular pieces of DNA.

Mitochondrial DNA Abnormalities

Mitochondria are intracellular organelles that generate energy via a series of respiratory chain complexes. They contain a unique circular chromosome that codes for 13 proteins, various RNAs, and several regulating enzymes. However, > 90% of mitochondrial proteins are coded by nuclear genes. Each cell has several hundred mitochondria in its cytoplasm.

Mitochondrial disease is due to mitochondrial DNA abnormalities (eg, deletions, duplications, mutations). High-energy tissues, such as muscle, heart, and brain, are particularly at risk, but hearing, pancreas, and liver are also at risk. Patterns of tissue involvement correlate with particular mitochondrial DNA changes, eg, chronic progressive external ophthalmoplegia; its variant, the multisystem **Kearns-Sayre syndrome** (chronic progressive external ophthalmoplegia, heart block, retinitis pigmentosa, CNS degeneration); **Pearson syndrome** (sideroblastic anemia, pancreatic insufficiency, and progressive liver disease that begins in the first few months of life and is frequently fatal in infants); **Leber's hereditary optic neuropathy** (a variable but often devastating bilateral visual loss that often occurs in teenagers and that is due to a point mutation in mitochondrial DNA); **MERRF** (*M*yoclonic *E*pilepsy, *R*agged *R*ed *F*ibers, dementia, ataxia, myopathy); and **MELAS** (*M*itochondrial *E*ncephalomyopathy, *L*actic *A*cidosis, *S*trokelike episodes). Mitochondrial pathology occurs in many common disorders (eg, large mitochondrial deletions in the cells of the basal ganglia of patients with Parkinson's disease, many types of

muscle disease, progressive accumulation of mitochondrial DNA deletions with aging).

Maternal inheritance characterizes abnormalities of mitochondrial DNA because all mitochondria are inherited via the egg. Thus, all offspring of an affected female are at risk of inheriting the abnormality, whereas no offspring of an affected male are at risk. Variability in clinical manifestations is the rule and may be due in part to variable mixtures of mutant and normal mitochondrial genomes (heteroplasmy) within cells and tissues.

Immunogenetics

The genetic basis of the immune system is very complex. Several genetic immune deficiencies involve various parts of the immune defense mechanism. The major histocompatibility complex is important in self-recognition. Inheritance of the major histocompatibility locus is autosomal recessive, so siblings have a 25% chance of having the same major histocompatibility locus. The histocompatibility locus is important for organ transplantation and plays a major role in tolerance for organ and bone marrow transplantation. Another component of the immune system involves RBC surface molecules, which cause immune reactions during blood transfusion. The ABO and Rh RBC antigen systems are particularly important in transfusion reactions and in maternal-fetal incompatibility.

A **clone** is derived from one somatic cell rather than from gamete fusion (egg and sperm). The cloning of mammals from single cells has recently been achieved. **Chimeras** are individuals with cells from genetically distinct sources (eg, transplantation, graft, embryonic fusion). By contrast, monozygotic twins are two

individuals who have developed after the fertilization of one egg by one sperm.

Forensic genetics

Newer molecular genetic techniques are very powerful in identifying a person's genetic makeup. Multiple genetic markers, each of which has a great deal of normal variability among persons, can be used to identify whether two persons are genetically related. These polymorphic markers are used in DNA techniques that identify whether one person is the offspring of another. Half of the markers should come from the father and half from the mother. Thus, analysis of blood samples and extracted DNA can determine whether the putative father and mother of a particular child are the biologic parents. Paternity can usually be determined precisely, except in the case of monozygotic twins, who have identical DNA markers.

The same polymorphic DNA markers can be used to identify a specimen and determine the person who is its source. Biopsies, pathologic specimens, blood smears, semen samples, tissues from under fingernails, and even a hair follicle can be used. With the polymerase chain reaction, the DNA of a single cell can be amplified, providing a sufficient quantity to determine the source of the DNA.

Genetic therapy

Therapy for genetic disease is often very similar to therapy for other types of disorders. Special diets can eliminate compounds that are toxic to patients, such as those in phenylketonuria and homocystinuria. Vitamins or other agents can improve a

biochemical pathway and thus reduce toxic levels of a compound. For example, folic acid reduces homocysteine levels in a person who carries the 5,10-methylene tetrahydrofolate reductase polymorphism. Therapy may involve replacing a deficiency or blocking an overactive pathway. A fetus can sometimes be treated by treating the mother (eg, corticosteroids for congenital virilizing adrenal hypoplasia) or by using in utero cellular therapy (eg, bone marrow transplantation). A newborn with a genetic disease may be a candidate for treatment with bone marrow or organ transplantation.

Gene therapy may involve the insertion of normal copies of a gene into the cells of persons with a specific genetic disease. Trials of somatic gene therapy have been undertaken for very severe genetic disorders (eg, adenosine deaminase deficiency). Germline gene therapy could involve the correction of an abnormality in the genes of sperm or egg and is considered an inappropriate way to deal with genetic diseases because of ethical issues, cost, lack of research in humans, lack of knowledge about whether changes would be maintained in the growing embryo, and the relative ease of treating somatically when needed. Gene therapy could also involve turning off genes as by antisense DNA.

Summary

Genetic Screening used in populations at risk for a particular genetic disorder.

1. **Heterozygote screening:** Screening a susceptible population (eg, Tay-Sachs disease in Ashkenazic Jews, sickle cell anemia in blacks.
2. **Presymptomatic genetic screening:** For persons with a family history of a dominantly inherited disorder (Huntington's disease).
3. **Prenatal diagnosis:** If maternal age > 35 yr.
4. **Newborn screening:** Screening for phenylketonuria, allows prophylaxis

Construction of a Family Pedigree (family tree)

To determine the genetic risk. It provides a view of illnesses within the family and facilitates analysis of inheritance patterns.

Genetic Counseling

Genetic counseling involves obtaining a thorough family history and addressing the family's concerns and questions.

What are genetic disorders?

A genetic disorder is a disease caused by abnormalities in an individual's genetic material (genome). Four different types of genetic disorders: (1) single-gene, (2) multifactorial, (3) chromosomal, and (4) mitochondrial.

(1) Single-gene: It is caused by mutations that occur in the DNA sequence of one gene, so that its protein product lakes its normal function.

Inheritance Of Single-Gene Defects

Pedigrees are based on the phenotype (observable features). With molecular studies, the genotype can also be determined and recorded.

- **Autosomal dominant**

A person need to have only one abnormal allele of a gene.

Expressivity and penetrance: A gene's effects may be influenced by the environment and by other genes that may alter phenotypic expression

Pleiotropy: A single-gene defect may produce multiple anomalies in different organ systems. For example, breakable bones, deafness, bluish whites of the eyes, dysplastic teeth, hypermobile joints, and heart valve abnormalities may occur in osteogenesis imperfecta.

Sex-limited inheritance: A trait that appears in only one sex is called sex-limited.

Mutations: An autosomal dominant pedigree begins with a fresh mutation.

- **Autosomal recessive:** A person must have two copies of an abnormal allele to develop an autosomal recessive disorder. (eg, sickle cell anemia).

- **X-linked dominant**

• Affected males transmit the trait to all of their daughters but not their sons.
• Affected heterozygous females transmit the condition to 1/2 of their children, regardless of sex.
• Affected homozygous females transmit the trait to all of their children.
• Twice as many affected females as males will have the disorder unless it is lethal in males.

- **X-linked recessive** In general, the following rules of inheritance apply:

• Nearly all affected persons are male.
• If the trait is transmitted through the heterozygous mother, she is usually phenotypically normal
• The trait may represent a new mutation in the affected male.
• An affected male never transmits the trait to his sons.
• All daughters of an affected male will be carriers.
• The carrier female transmits the trait to 1/2 of her sons.
• No daughters of a carrier female will show the trait, but 1/2 will be carriers (unless they also inherit the trait from their father, such as color blindness).

- **Codominant**

In codominant inheritance, both alleles are expressed in a heterozygous person, as the blood group antigens (eg, AB, MN).

(2) Multifactorial: This type is caused by a combination of environmental factors and mutations in multiple genes. For example, different genes that influence breast cancer susceptibility.

Nontraditional inheritance

1. **Mosaicism:** Mosaicism is the presence of 2 cell lines differing in genotype or karyotype but derived from one zygote.

2. **Uniparental disomy:** Uniparental disomy occurs when two chromosomes of a pair are inherited from only one parent. This is very rare and is thought to involve trisomy rescue

3. **Triplet repeat, unstable mutations:** A triplet repeat is an unusual type of mutation in which a triplet of nucleotides increases in number within a gene.

4. **Anticipation:** Anticipation occurs when a disorder has an earlier age of onset and severity of expression in each successive generation. It may occur because a parent is a mosaic, and the child has the full mutation in all cells. Triplet repeat expansion may demonstrate anticipation when the number of repeats increases with each generation.

(3) Chromosomal: Abnormalities in chromosome structure as missing or extra copies or gross breaks and rejoinings (translocations), can result in genetic disease.

Congenital anomalies

Congenital anomalies may be inherited or sporadic, isolated or multiple, apparent or hidden, gross or microscopic.

Congenital heart disease

Anatomic defects of the heart and great vessels produced at various stages of fetal development and present at birth.

Cleft lip and **cleft palate** are the most common 1st arch defects,

Spinal abnormalities Congenital torticollis

Hip, leg, and foot abnormalities: Congenital dislocation of the hip Femoral torsion Knee dislocation Bowing and twisting (torsion) of the tibia

Abnormalities of bone: Achondrodystrophies and osteogenesis imperfecta

(4) Mitochondrial - This relatively rare type of genetic disorder is caused by mutations in the nonchromosomal DNA of mitochondria.

CHAPTER 15



ANCER

Introduction

Cellular proliferation is under genetic control, and if somatic mutation creates a variant that proliferates faster, the mutated cells will tend to take over the organism. Thus people have a natural tendency to have tumors.

The evolution from a normal somatic cell to a malignant tumor takes place within the life of an individual, and has to start afresh with each new individual. But an organism with a good anti-tumor mechanism transmits it to its offspring, where it continues to evolve. Potential tumor cells are either repaired and brought back into line, or made to kill themselves (apoptosis). No single mutation can escape these mechanisms and convert a normal cell into a malignant one. Long ago, studies of the age-dependence of cancer suggested that average 6-7 successive mutations are needed to convert a normal cell into an invasive carcinoma. In other words. The chance of a single cell undergoing six independent mutations is negligible, suggesting that cancer should be rare.

Not surprisingly, carcinogenic mutations usually affect the genes that control the birth (cell cycling) or death (apoptosis) of cells. Two broad categories of genes can be distinguished, **Oncogenes and Tumor suppressor (TS) genes.** By analogy with a car, the oncogenes as the accelerator and the tumor suppressor genes as the brake. Jamming the accelerator on (a

dominant gain of function of an oncogene) or having all the brakes fail (a recessive loss of function of a TS gene) will make the car run out of control.

Oncogenes

These are genes whose normal activity promotes cell proliferation. Gain of function mutations in tumor cells create forms that are excessively or inappropriately active. The non-mutant versions are properly called proto-oncogenes.

Viral oncogenes

For many years it has been known that some animal leukemias, lymphomas and cancers are caused by viruses. Tumor viruses fall into three broad classes:

- **DNA viruses** integration of the viral genome implants the transcriptional activation or replication signals of the virus into the host genome and triggers cell proliferation. Some of the viral genes involved have been identified,
- **Retroviruses** have a genome of RNA. They replicate via a DNA intermediate, which is made using a viral reverse transcriptase. These viruses do not normally kill the host cell (HIV is an exception), and only transform it. The genome of a typical retrovirus consists of three genes, gag, pol and env.
- **Acute transforming retroviruses** are retrovirus which, unlike normal retroviruses, transform the host cell rapidly and with high efficiency. Their genomes include an additional gene, the oncogene (50 different oncogenes.).

Proto-oncogenes

It quickly became clear that normal cells had counterparts of the retroviral oncogenes, and in fact that v-onc (viral oncogenes) genes were transduced cellular genes. With a few exceptions, the v-onc gene products differ from their c-onc (proto-oncogene) counterparts by amino acid substitutions or truncations, which serve to activate the proto-oncogene.

Functional understanding of oncogenes began with the discovery in 1983 that the viral oncogene v-sis was derived from the normal cellular platelet-derived growth factor B (PDGFB) gene. Uncontrolled over-expression of a growth factor would be an obvious cause of cellular hyperproliferation. The roles of many cellular oncogenes (proto-oncogenes) have now been elucidated. Five broad classes can be distinguished:

- secreted growth factors (e.g. SIS);
- cell surface receptors (e.g. ERBB, FMS);
- components of intracellular signal transduction systems (e.g. the RAS family, ABL);
- DNA-binding nuclear proteins, including transcription factors (e.g. MYC, JUN);
- components of the network of cyclins, cyclin-dependent kinases and kinase inhibitors that govern progress through the cell cycle (e.g. MDM2).

Activation of proto-oncogenes

Activation involves a gain of function. This can be quantitative (an increase in the production of normal protein) or qualitative (production of a modified protein as a result of a mutation, or production of a novel product from a chimeric gene created by a

chromosomal rearrangement). These changes are dominant and normally affect only a single allele of the gene.

Inheritance

Activating mutations in oncogenes occur in somatic cells. Constitutional mutations (occur in gametes) would probably be lethal to the new offspring. There is one exception: specific activating point mutations in the RET oncogene cause multiple endocrine neoplasia or familial thyroid cancer, and sometimes these mutations are inherited. These RET mutations must affect the behavior of only very specific cells in very special circumstances.

Note however that nonactivating mutations in proto-oncogenes may be inherited constitutionally, if their effect is unrelated to cancer. For example, inherited mutations that inactivate the KIT oncogene produce piebaldism, while inherited loss-of-function mutations in RET predispose to Hirschsprung's disease.

Activation by amplification

Many cancer cells contain multiple copies of structurally normal oncogenes. Breast cancers often amplify ERBB2, MYC, and NMYC. Hundreds of extra copies may be present. They can exist as small separate chromosomes (double minutes) or as insertions within the normal chromosomes.

Activation by point mutations

The H-RAS1 gene is one of a family of ras genes that encode proteins. Specific point mutations in RAS genes lead to amino acid substitutions that decrease the activity of the RAS protein leading to excessive cellular division. This is frequently found in

cells of many tumors including colon, lung, breast and bladder cancers.

Activation by chromosomal translocations

Tumor cells typically have grossly abnormal karyotypes, with multiple extra and missing chromosomes, many translocations and so on. Most of these changes are random, and reflect a general genomic instability which is a normal part of carcinogenesis. Over 150 different tumor-specific breakpoints have now been recognized and they reveal an important common mechanism in tumorigenesis.

The best-known tumor-specific rearrangement produces the Philadelphia (Ph^1) chromosome, a very small acrocentric chromosome seen in 90% of patients with chronic myeloid leukemia. This chromosome turns out to be produced by a balanced reciprocal 9;22 translocation. The breakpoint on chromosome 9 is within an intron of the ABL oncogene. The translocation joins most of the ABL genomic sequence onto a gene called BCR (breakpoint cluster region) on chromosome 22, creating a novel fusion gene. This chimeric gene is expressed to produce a tyrosine kinase related to the ABL product but with abnormal transforming properties.

Activation by transposition to an active chromatin

Burkitt's lymphoma is a childhood tumor common in malarial regions of Central Africa. A characteristic chromosomal translocation, t(8;14)(q24;q32) is seen in 75-85% of patients. The remainder have t(2;8)-(p12;q24) or t(8;22)(q24;q11). Each of these translocations puts the MYC oncogene close to an immunoglobulin locus, IGH at 14q32, IGK at 2p12 or IGL at

22q11. The Burkitt's lymphoma translocations do not create novel chimeric genes. Instead, they put the oncogene in an environment of chromatin that is actively transcribed in antibody-producing B-cells.

Tumor suppressor genes

There are evidence that tumorigenesis involves not only dominant activated oncogenes, but also recessive mutations (both alleles of gene must be inactivated) in other genes. These other genes are the tumor suppressor (TS) genes. Sometimes TS genes are called antioncogenes, but that is an unhelpful name because it wrongly implies that they are all specific antagonists or inhibitors of oncogenes. TS genes can have a variety of functions (1) prevent cell cycle progression, (2) steer deviant cells into apoptosis, and (3) keep the genome stable and mutation rates low by ensuring accurate replication, repair and segregation of the cell's DNA.

Retinoblastoma (Knudson's two-hit hypothesis)

Retinoblastoma is a rare, aggressive childhood tumor of the retina. 60% of cases are sporadic and unilateral; the other 40% are inherited as an imperfectly penetrant autosomal dominant trait, which was mapped to 13q14. In familial retinoblastoma bilateral tumors are common. In 1971 AG Knudson proposed that two successive mutations ('hits') were required to turn a normal cell into a tumor cell, and that in familial forms one of the hits was inherited.

Cavenee and colleagues sought evidence of somatic mutations at the RB1 locus in sporadic retinoblastoma by typing surgically removed tumor material with a series of markers from chromosome 13.

Methylation of tumor suppressor genes

Tumor supressor genes may be silenced by deletion (reflected in loss of heterozygosity) or by point mutations, but there is increasing evidence for a third mechanism - DNA methylation. Cytosines in CpG dinucleotides are liable to be methylated, and when the cytosine lies in a CpG island within the promoter region of a gene, methylation is often associated with lack of expression of the gene. Many tumors show genome-wide disturbances of the normal methylation pattern. More specifically, CpG island methylation has been demonstrated for several tumor suppressor genes in a variety of cancers.

Control of the cell cycle

Any cell at any time has three choices of behavior: it can remain static, it can divide or it can die (apoptosis). Some cells also have the option of differentiating. Cells select one of these options in response to internal and external signals(stimuli). Oncogenes and tumor suppressor genes play key roles in generating and interpreting these signals.

Life would be very simple if the signal and response were connected by a single linear pathway, but this seems never to be the case. Rather, multiple branching, overlapping and partially redundant pathways control the behavior of the cell. Probably such complicated networks are necessary to confer stability and resilience on the extraordinarily complex machinery of a cell.

shows part of the cell cycle control system that involves the products of three key genes, RB1, TP53 and CDKN2A. One way or another, tumor cells must inactivate this control system - probably they need to inactivate both the RB1 and TP53 arms of

the system. Thus these three tumor suppressor genes are central players in carcinogenesis, and are among the most commonly altered genes in tumor cells. Each also has a role in inherited cancers.

Function of pRb, the RB1 gene product

The RB1 gene is widely expressed, encoding a 110-kd nuclear protein, (pRb). In normal cells pRb is inactivated by phosphorylation and activated by dephosphorylation. Active (dephosphorylated) pRb binds and inactivates the cellular transcription factor E2F1, function of which is required for cell cycle progression. The G_1-S checkpoint seems to be the most crucial in the cell cycle; 2-4 hours before a cell enters S-phase, pRb is phosphorylated (inactive). This leads to activation of E2F1 and allows the cell to proceed to S phase. Phosphorylation is governed by a cascade of cyclins, cyclin-dependent kinases and cyclin kinase inhibitors.

RB1 (retinoblastoma)gene mutations produce sporadic or inherited retinoblastoma. It is not clear why constitutional mutation of a gene which is essential to control cell cycle result specifically in retinoblastoma and a small number of other tumors, principally osteosarcomas.

The product of the MDM2 oncogene (which is amplified in many sarcomas) binds and inhibits pRb, thus favoring cell cycle progression. Several viral oncoproteins (adenovirus E1A, SV40-T antigen, human papillomavirus E7 protein) also bind and sequester or degrade pRb.

Function of p53, the TP53 gene product

p53 was first described in 1979 as a protein found in SV40-transformed cells, where it associated with the T-antigen. Later, the TP53 gene which encodes p53 appeared as a dominant transforming gene, and so was classed as an oncogene. Subsequently it transpired that while p53 from some tumor cells was oncogenic, p53 from normal cells positively suppressed tumorigenesis.

Loss or mutation of TP53 is probably the commonest single genetic change in cancer. This reflects the central importance of p53, which has several functions in the cell. (1) It acts as a transcription factor. It binds DNA and can activate transcription of reporter genes placed downstream of a p53 binding site. (2) it can stop replication of the cells which have damaged DNA so it is called 'the guardian of the genome'. Normal cells with damaged DNA arrest at the G_1-S cell cycle checkpoint until the damage is repaired, but cells that lack p53 or contain a mutant form do not arrest at G_1. Replication of damaged DNA presumably leads to random genetic changes, some of which are oncogenic, similar to cells with a defective mismatch repair system.

Probably p53 has a crucial role in cell death. In response to oncogenic stimuli, cells undergo apoptosis (programmed cell death). Apoptosis has come to occupy a central place in our understanding of the cancer. Tumor cells lacking p53 do not undergo apoptosis, and so escape the control.

Loss of heterozygosity assays confirmed the status of TP53 as a tumor suppressor gene. TP53 maps to 17p12, and this is one of the commonest regions of loss of heterozygosity in a wide range of tumors. Tumors that have not lost TP53 very often

have mutated versions of it. To complete the picture of TP53 as a TS gene, constitutional mutations in TP53 are found in families with the dominantly inherited Li-Fraumeni syndrome. Affected family members suffer multiple primary tumors, typically including soft tissue sarcomas, osteosarcomas, tumors of the breast, brain and adrenal cortex, and leukemia.

Checkpoint of DNA damage

Normal cells with unrepaired DNA damage do not enter mitosis. Among the mammalian genes involved in the damage checkpoint are ATM and maybe ATR, BRCA1 and BRCA2.

ATM is the gene responsible for ataxia telangiectasia a rare recessive combination of cerebellar ataxia, telangiectasia (dilation of blood vessels in the conjunctiva and eyeballs), immunodeficiency, growth retardation and sexual immaturity. ATM patients have a strong predisposition to cancer. Homozygotes usually die of malignant disease before age 25, and there have been suggestions that heterozygotes have a raised risk of cancer - for example a 3.9-fold increased risk of breast cancer among women.

ATR (A related protein)inhibits the p53 response to DNA damage.

BRCA1 and BRCA2 (breast cancer genes), may also form part of the checkpoint mechanism.

In all cases, if cells replicate with damaged DNA, there is increased risk of mutations, DNA breaks or crosslinks which predispose to the chromosome deletions, translocations and inversions that are such a common feature of tumor cells.

Telomeres, and cancer cells

The ends of human chromosomes are protected by a repeat sequence of $(TTAGGG)_n$, (telomeres) that is maintained by a telomerase enzyme system,. Telomerase is present in the human germ line but is absent in most somatic tissues, and so telomere length declines with time in normal somatic cells. This phenomenon which may contribute to the 'mitotic clock' that limits the number of divisions for any cell. There has been much excitement over the discovery that 90% of human primary tumors possess telomerase activity. Maybe this is the key to their immortality, and may be an anti-telomerase agent would limit their mitotic potential.

Summary

Malignancy is defined as uncontrollable cellular proliferation which is under genetic control. and if somatic mutation creates a variant that proliferates faster, the mutated cells will tend to have tumors. 6-7 successive mutations are needed to convert a normal cell into malignant cells. Any cell at any time has three choices of behavior: it can remain static, it can divide or it can die (apoptosis). Cells select one of these options in response to internal and external signals(stimuli). Oncogenes and tumor suppressor genes play key roles in generating and interpreting these signals.

Oncogenes

These are genes whose normal activity promotes cell proliferation (proto-oncogenes). Gain of function mutations in tumor cells create forms that are excessively or inappropriately active.

Viral oncogenes

Tumor viruses fall into three broad classes:

- **DNA viruses** integration into the host genome triggers cell proliferation.
- **Retroviruses** have a genome of RNA. They replicate via a DNA intermediate, which is made using a viral reverse transcriptase.
- **Acute transforming retroviruses** are retrovirus transform the host cell rapidly and with high efficiency.

Activation of proto-oncogenes

This can be quantitative (an increase in the production of normal protein) or qualitative (production of a modified protein). These changes are dominant and normally affect only a single allele of the gene.

1-Activation by amplification

Many cancer cells contain multiple copies of structurally normal oncogenes. Breast cancers often amplify ERBB2, MYC, and NMYC.

2-Activation by point mutations

The H-RAS1 gene is one of a ras genes family. Point mutations in RAS genes lead to amino acid substitutions and so, excessive cellular division..

3-Activation by chromosomal translocations

Tumor cells typically have abnormal karyotypes, with multiple extra and missing chromosomes, many translocations which reflect a general

genomic instability (Philadelphia (Ph^1) chromosome, in 90% of patients with chronic myeloid leukemia).

4-Activation by transposition to an active chromatin

Burkitt's lymphoma translocations do not create novel chimeric genes. Instead, they put oncogene in an environment of chromatin that is actively transcribed in B-cells.

5-Inherited

Activating mutations in oncogenes occur in somatic cells. Constitutional mutations (occur in gametes) would probably be lethal to the new offspring

Tumor suppressor genes

TS genes can have a variety of functions (1) prevent cell cycle progression, (2) steer deviant cells into apoptosis, and (3) keep the genome stable and mutation rates low by ensuring accurate replication, repair and segregation.

Function of pRb, the RB1 gene product

Phosphorylated (inactive). pRb leads to activation of E2F1 and allows the cell to proceed to S phase. RB1 (retinoblastoma)gene mutations produce sporadic or inherited retinoblastoma. and a small number of other tumors, principally osteosarcomas .

Function of p53, the TP53 gene product

Loss or mutation of TP53 is the commonest single genetic change in cancer. p53, has several functions (1) acts as a transcription factor (2) can stop damaged DNA replication 'the guardian of the genome'.(3) has a crucial role in programmed cell death. (apoptosis).

Checkpoint of DNA damage

Normal cells with unrepaired DNA damage do not enter mitosis. If cells replicate with damaged DNA, there is increased risk of mutations, DNA breaks or crosslinks which predispose to the chromosome deletions, translocations and inversions that are such a common feature of tumor cells.

Telomeres, and cancer cells

The ends of human chromosomes are protected by a repeat sequence of $(TTAGGG)_n$, (telomeres) that is maintained by a telomerase enzyme system,. Telomerase is present in the human germ line but is absent in most somatic tissues, and so telomere length declines with time in normal somatic cells. 90% of human primary tumors possess telomerase activity.

CHAPTER 16



GENOME PROJECTS

Introduction

For the first time, we will know the sequence of our DNA. Sequencing of our DNA will be just the beginning of a huge effort to understand exactly how this sequence can specify a person, and how the DNA of other organisms is related to us and to their biologies.

The Human Genome Project is an international effort, the primary aim of HGP is to deliver the complete nucleotide sequence of our DNA. The human mtDNA sequence was established in 1981, so the genome in question is the nuclear genome. It has been paralleled by many other genome projects which seek to sequence the DNA of a variety of model organisms. Some genome projects have already been completed including the Human Genome Project.

History

The first workshop of the Human Genome Project held in Alta, in December 1984. Principal conclusion was that methods were incapable of measuring mutations with sufficient sensitivity unless an enormously large, complex and expensive program was undertaken. A subsequent report on Technologies for Detecting Heritable Mutations in Human Beings sparked the idea for a dedicated human genome project by the DOE.

Although such technologies had been in operation for about a decade, the efforts of individual laboratories to try to clone and characterize one gene at a time were considered to be wasteful of scientists' time and research resources.

In March 1986, it sponsored an international meeting in Santa Fe, New Mexico, to assess the desirability and feasibility of ordering and sequencing DNA clones representing the entire human genome. Virtually all participants concluded that such a project was feasible and would be an oustanding achievement in biology.

In 1988, two additional reports from the US Office of Technology Assessment and National Research Council appeared.

In the same year, the US congress officially gave approval to a 15-year US human genome project commencing in 1991. The required funding was estimated to be about $3 billion.

Organization

The USA Human Genome Project remains the major contributor to international research in this area, but several other countries quickly developed their own Human Genome Projects. Centers in the UK and France have made major contributions and large programs are also underway in some other countries, notably Germany and Japan.

In order to coordinate the different national efforts, the Human Genome Organization (HUGO) was established in Human Genome Project is concentrated in a few very large genome centers, but interacting with them is a worldwide

network of small laboratories, mostly attempting to map and identify disease genes.

Goals

- To acquire fundamental information concerning our genetic
- understanding the role of human genes in health and diseases.
- To achieving high-resolution genetic maps
- To construct a framework for high-resolution physical maps.
- To complete the sequence of the human genome.
- **Medical and scientific benefits**
  1. It will provide more comprehensive prenatal and presymptomatic diagnoses of disorders in individuals judged to be at risk of carrying a disease gene.
  2. It will alter the current approach to medical care, from one of treating advanced disease to preventing disease based on the identification of individual risk
  3. The information on gene structure will also be used to explore how individual genes function and how they are regulated.
  4. It will provide the needed explanations for biological processes in human and all organisms.
  5. It would provide a framework for developing new therapies for diseases, in addition to simple gene therapy approaches.

Genetic mapping

Classical genetic maps for experimental organisms such as **Drosophila** and **mouse** are based on gene studies. Unlike the experimental organisms, the human genetic map was never going to be based on genes because the frequency of mating between two individuals suffering from different genetic disorders is extremely small.

The first human genetic map

The desirability of a complete linkage map of the human genome was clear. In 1987, after a huge effort, the first such map was published based on the use of 403 polymorphic loci, including 393 RFLP markers. The average spacing between the markers (>10 cM) was still considerable, and, more significantly, RFLP markers are not very informative and are difficult to type.

High-resolution genetic maps (second-generation)

Hypervariable minisatellite map polymorphisms are highly polymorphic, but their applicability to genome-wide maps is limited because they are mostly restricted to chromosomal regions near the telomeres.

Microsatellite markers map have the advantage of being abundant, dispersed throughout the genome, highly informative and easy to type. Researchers at the Généthon laboratory in France were quickly able to provide a second-generation linkage map of the human genome. With an increasing numbers of microsatellite markers, the resolution. a genetic map became 1 cM. After this time the major effort switched to the construction of high-resolution physical maps.

Physical mapping

The physical map of the human genome will consist of 24 maps, one for each chromosome.

The first physical map of the human genome was obtained more than 40 years ago when cytogenetic banding techniques were used not only to distinguish the different chromosomes, but also to provide discrimination of different subchromosomal regions (human karyogram). Although the resolution is coarse (an average sized chromosome band in a 550-band preparation contains ~6 Mb of DNA), it has been very useful as a framework for ordering the locations of human DNA sequences by chromosome in situ hybridization techniques.

Second physical map have been obtained by mapping natural chromosome breakpoints (using translocation and deletion hybrids), or by mapping artificial chromosome breakpoints using radiation hybrids, but the resolution achieved can be quite limited. Such maps have, however, been useful frameworks for mapping genes (transcript maps).

Large-scale restriction maps have also been generated, such as the NotI restriction map of 21q. However, the most important maps are clone contig maps because these are the immediate templates for DNA sequencing.

YAC clone contig maps

First-generation physical map of the human genome: A complete clone contig map of a chromosome would comprise all the DNA without any gaps (contig originated as a shortened form of the word contiguous;). Yeast artificial chromosome (YAC) clones have been particularly useful in generating first-generation

physical maps of human chromosomes because they can clone large sequence of DNA.

Significant contig maps for individual human chromosomes were first reported in 1992 for chromosome 21 and the Y chromosome and, subsequently, a first-generation clone contig map of the human genome was reported by workers at the CEPH lab in Paris. An updated YAC contig map, covering perhaps 75% of the human genome and consisting of 225 contigs with an average size of 10 Mb, was subsequently published by the same group.

BAC/PAC clone contig maps

The utility of YAC contig maps is limited because YAC inserts are not representations of the original starting DNA; many YAC clones are chimeric or have internal deletions.

Second-generation clone contig maps have relied on bacterial artificial chromosomes (BACs) and P1 artificial chromosomes (PACs). Although the insert sizes of these clones (typically 70-250 kb) are much smaller than that of YACs, they show great stability, making them more representations of the original DNA.

Human genome sequence

There have been no significant changes in the basic sequencing technology; the dideoxy sequencing approach invented by Fred Sanger and his colleagues at Cambridge, UK, more than 20 years ago is still used. Instead, efficiency gains have been made through the use of automated fluorescence-based systems and capillary gel electrophoresis.

While the first few years of the Human Genome Project were devoted to producing high-resolution genetic and physical maps, large-scale human genome sequencing is now very much underway and 10% of the human genome had been sequenced by May 1999.

The Human Genome Diversity Project

The Human Genome Project was conceived as a project to obtain the nucleotide sequence of a collection of cloned human DNA fragments. Information of human genetic showed diversity which has four major contexts:

- **Human evolution**. The information should be of help in anthropological and historical research in tracking human origins, prehistoric population movements and social structure.
- **Identification of common disease genes** Common diseases are multifactorial, and it can be frustratingly difficult to identify the underlying genes
- **Identification of factors which confer susceptibility to or protection from genetic diseases..** Genetic differences between human populations can make some populations more susceptible to particular diseases while others are comparatively protected.
- **Forensic anthropology**. The accuracy of DNA fingerprinting, a widely used tool in forensic science, is dependent on knowing how the DNA markers detected in fingerprinting vary from one population to the next.

Single nucleotide polymorphism (SNP) maps

Human genetic maps based on single nucleotide polymorphisms (SNPs) are very frequent (about 1 per kb) and typing is easily automated

Other genome projects

Prokaryotic genomes are typically small (often only one or a few Mb) and are therefore amenable to comparatively rapid sequencing. By 1999 the genomes of a total of 75 different prokaryotes were being sequenced or had already been sequenced. The first to be completed (in 1995) was the 1.83 Mb genome of Haemophilus influenzae.

Pathogenic organisms The UK Wellcome Trust has been a major supporter of genome projects for microbial pathogens. They include Helicobacter pylori (associated with peptic ulcers) and Chlamydia penumoniae (associated with respiratory disease and also with coronary heart disease). reponema pallidum (the causative agent of syphilis) and Rickettsia prowazekii (the causative agent of typhus). The new information can be expected to lead to more sensitive diagnostic tools and new targets for establishing drugs and vaccines.

Life in the post-genome (sequencing) era

Once the sequence of the human genome is known, what difference will it make? Certainly, there will be a huge boost to basic research as we grapple with the fundamental biological question of how our genome is interpreted to specify a person.

post-genome era, an accurate genetic testing will become widely available, not just for genetic disorders, but also in terms

of genetic susceptibility to a variety of different conditions, including infectious diseases. Improved treatments can also be expected. Gene therapy approaches may prove technically difficult, but the new infomation will undoubtedly assist the development of novel therapies.

Comparative genomics

involves analysis of two or more genomes to identify the extent of similarity of various features, or large-scale screening of a genome to identify sequences present in another genome. The examples below are merely meant to be illustrative of some of the applications.

- **Evolutionary relationships**. One early application involved comparison of archaeal genomes with eubacterial and eukaryotic genomes to infer evolutionary relatedness.
- **Identification of regulatory elements**. We have very limited information about regulatory elements in complex eukaryotic genomes. By referring to databases of known regulatory element sequences, computer programs can inspect new genomic sequences for the presence of regulatory elements, but the efficiency is very low.
- **Gene identification**. Electronic screening of EST databases can identify homologs of biologically interesting genes in other species.

Functional genomics

Similar types of approaches will also permit extensive investigation of human gene function and dissection of complex regulatory pathways. Once all human genes are known, we can know all the products. Now as we explore RNA and polypeptide

expression products on a global scale, new terms are being coined: the **transcriptosome** (collection of RNA transcripts in a cell) and the proteome (collection of polypeptides/proteins expressed in a cell).

The new science of **proteomics** is devoted to the study of global changes in protein expression and the systematic study of protein-protein interactions.

Safeguards

Any major scientific advance carries with it the fear of exploitation. The Human Genome Project is not an exception, and the perceived benefits of the project can also have a downside. For example,

Once we know all the human genes and detect the associated mutations, there will be benefit in prevention of disease to those individuals who can carry mutations for that disease. However, the same information can also be used against such individuals by insurance companies. For example, there is the very real prospect of insurance companies insisting on genetic screening tests for the presence of genes that confer susceptibility to common disorders, such as diabetes, cardiovascular disease, cancers and various mental disorders. Perfectly healthy individuals who happen to be identified as carrying such disease-associated alleles may then be refused life or medical insurance.

A fundamental ethical principle in all genetic counseling and genetic testing is that genetic information should be generated only in response to an explicit request from a fully informed adult patient.

Another troublesome area is **eugenics**, (the application of selective breeding or other genetic techniques) to 'improve' human qualities. In recognition of the above problems, the US Human Genome Project has devoted considerable resources to support research into the ethical, legal and social impact of the project.

Summary

The Human Genome Project is an international effort, aim to deliver the complete nucleotide sequence of our DNA. Some genome projects have already been completed including the Human Genome Project.

Goals

- To acquire fundamental information concerning our genetic
- understanding the role of human genes in health and diseases.
- To achieving high-resolution genetic maps
- To construct a framework for high-resolution physical maps.
- To complete the sequence of the human genome.
- **Medical and scientific benefits**

1. prenatal and presymptomatic diagnoses of gene disorders
2. preventing disease based on the identification of individual risk
3. explore how individual genes function and how they are regulated.
4. explain biological processes in human and all organisms.
5. develop new therapies for diseases, and gene therapy.

Genetic mapping: Classical genetic maps for experimental organisms such as Drosophila and mouse are based on gene studies based on genes study

The first human genetic map

It was based on the use of 403 polymorphic loci, including 393 RFLP markers.

High-resolution genetic maps (second-generation).

Hypervariable minisatellite map: applicability to genome-wide maps is limited because they are mostly restricted to chromosomal regions near the telomeres.

Microsatellite markers map abundant, dispersed throughout the genome, highly informative and easy to type.

Physical mapping The physical map of the human genome will consist of 24 maps, one for each chromosome.

The first physical map: It uses cytogenetic banding to distinguish the different chromosomes, and to provide subchromosomal regions (human karyogram).

Second physical map by mapping natural chromosome breakpoints or by mapping artificial chromosome breakpoints

Large-scale restriction maps have also been generated, such as the NotI restriction map of 21q.

First-generation physical map of the human genome: Yeast artificial chromosome (YAC) clones have been particularly useful in generating first-generation physical maps of human chromosomes because they can clone large sequence of DNA.

Second-generation clone contig maps have relied on bacterial artificial chromosomes (BACs) and P1 artificial chromosomes (PACs).

The Human Genome Diversity Project

The Human Genome Project was conceived as a project to obtain the nucleotide sequence of a collection of cloned human DNA fragments. Information of human genetic showed diversity which has four major contexts:

- **Human evolution**. in anthropological and historical research in tracking human origins, prehistoric population movements and social structure.
- **Identification of common disease genes** Common diseases are multifactorial, and it can be frustratingly difficult to identify the underlying genes
- **Identification of factors which confer susceptibility to or protection from genetic diseases.**
- **Forensic anthropology**. The accuracy of DNA fingerprinting, a widely used tool in forensic science,.

Comparative genomics involves analysis of two or more genomes to identify the extent of similarity of various features: (1)Evolutionary relationships. (2) Identification of regulatory elements. (3)Gene identification.

Functional genomics

The new science of **proteomics** is devoted to the study of changes in protein expression and the systematic study of protein-protein interactions.

CHAPTER 17

GENE THERAPY

Introduction

Once a human disease gene has been characterized, molecular genetic tools can be used to dissect gene function and explore the biological processes involved in the normal and pathogenic states. The resulting information can be used to design novel therapies using conventional drug-based approaches. In addition, molecular genetic technologies have recently provided a variety of novel therapeutic approaches that can be categorized into two broad groups, depending on whether the therapeutic agent is a gene product/vaccine or genetic material.

Gene therapy

The term gene therapy describes any procedure intended to treat a disease by genetically modifying the cells of a patient. The genetic material may be (1) transferred directly into cells within a patient (in vivo gene therapy), (2) cells may be removed from the patient and the genetic material inserted into them in vitro, then transplanting the modified cells into the patient (ex vivo gene therapy). Because the molecular basis of diseases can vary widely, some gene therapy strategies are particularly suited to certain types of disorder, and some to others. Major disease classes include:

- infectious diseases (as a result of infection by a virus or bacterial pathogen);

- cancers (inappropriate continuation of cell division and cell proliferation as a result of activation of an oncogene or inactivation of a tumor suppressor gene or an apoptosis gene);
- inherited disorders (genetic deficiency of an individual gene product or genetically determined inappropriate expression of a gene);
- immune system disorders (includes allergies, inflammations and also autoimmune diseases, in which body cells are inappropriately destroyed by immune system cells).

A major motivation for gene therapy has been the need to develop novel treatments for diseases for which there is no effective conventional treatment. Gene therapy has the potential to treat all of the above classes of disorder. Depending on the basis of pathogenesis, different gene therapy strategies can be considered as follows:

- Classical gene therapy. The rationale of this type of approach is to deliver genes to appropriate target cells with the aim of obtaining optimal expression of the introduced genes. Once inside the desired cells in the patient, the expressed genes are intended to do one of the following:

  a. Produce a product that the patient lacks.
  b. Kill diseased cells directly, e.g. by producing a toxin which kills the cells.
  c. Activate cells of the immune system so as to aid killing of diseased cells.

- ♦ Non-classical gene therapy. The idea here is to inhibit the expression of genes associated with the pathogenesis, or to correct a genetic defect and so restore normal gene expression.

Current gene therapy is exclusively somatic gene therapy, the introduction of genes into somatic cells of an affected individual. The prospect of human germline gene therapy raises a number of ethical concerns, and is currently not sanctioned.

Genetically engineered hormones

Recombinant human insulin was first marketed in 1982 and, subsequently, a number of other cloned human gene products of medical interest have been produced commercially. Treatment with the products of cloned genes is not free from risks, however. For example, patients who completely lack a normal product may mount a vigorous immune response to the administered pharmaceutical product as in the case of some patients with severe hemophilia A who have been treated with recombinant factor VIII.

Genetically engineered antibody

Antibodies are natural therapeutic agents which are produced by B lymphocytes. In each B-cell precursor, a cell-specific rearrangement of antibody gene segments occurs so that individual B cells produce different antibodies. Additional diversity is provided by other mechanisms, including frequent somatic mutation events.

Artificially produced therapeutic antibodies are designed to be monospecific (they recognize a single type of antigenic site) and can recognize specific disease-associated antigens,

leading to killing of the disease cells. Notable targets for such therapy are cancers (especially lymphomas and leukemias); infectious disease (using antibodies raised against antigens of the relevant pathogen); and autoimmune disorders (where antibodies recognize inappropriately expressed host cell antigens).

Genetically engineered vaccines

Recombinant DNA technology is also being applied to the construction of novel vaccines.

Classical gene therapy

An essential component of classical gene therapy is that cloned genes have to be introduced and expressed in the cells of a patient in order to overcome the disease. Two major general approaches are used in the transfer of genes for gene therapy: transfer of genes into patient cells outside of the body (ex vivo) or inside the body (in vivo).

Ex vivo gene transfer

the cells are collected initially from the patient to be treated by new intact gene and grown in culture before being reintroduced into the same individual. Clearly, this approach is only applicable to tissues that can be removed from the body, altered genetically and returned to the patient where they will engraft and survive for a long period of time (e.g. cells of the hematopoietic system and skin cells).

In vivo gene transfer

Here the cloned genes are transferred directly into the tissues of the patient. This may be the only possible option in tissues where individual cells cannot be cultured in vitro in sufficient numbers

(e.g. brain cells) and/or where cultured cells cannot be re-implanted efficiently in patients. Liposomes and certain viral vectors are increasingly being employed for this purpose.

Principles of gene transfer

Minigene.

Classical gene therapies normally require efficient transfer of cloned genes into disease cells so that the introduced genes are expressed at suitably high levels. an artificial minigene (little form of the gene) may be used.

Genes integrated into chromosomes

As progeny cells contain the introduced genes, long-term gene therapy at least offers the opportunity for selecting cells where integration has been successful, amplifying them in cell culture and then checking the phenotypes for any obvious evidence of neoplastic transformation, prior to transferring the cells back into the patient.

Nonintegrated genes

Some gene transfer systems are designed to insert genes into cells where they remain as extrachromosomal elements and may be expressed at high levels. For example, cancer gene therapies often involve transfer and expression of genes into cancer cells with a view to killing the cells. Once the malignancy has been eliminated, the therapeutic gene may no longer be needed.

Vectors

The method chosen for gene transfer depends on the nature of the target tissue and whether transfer is to cultured cells ex vivo or to

the cells of the patient in vivo. No one gene transfer system is ideal; each has its limitations and advantages.

Oncoretroviral vectors

Retroviruses are RNA viruses which possess a reverse transcriptase function, enabling them to synthesize a complementary DNA Retroviruses are very efficient at transferring DNA into cells, and the integrated DNA can be stably propagated, offering the possibility of a permanent cure for a disease. Because of these properties, retroviruses were considered the most promising vehicles for gene delivery and currently about 60% of all approved clinical protocols utilize retroviral vectors.

Adenovirus vectors

Adenoviruses are DNA viruses that produce infections of the upper respiratory tract and have a natural tropism for respiratory epithelium, the cornea and the gastrointestinal tract. Adenovirus vectors have been the second most popular delivery system in gene therapy (with extensive applications in gene therapy for cystic fibrosis and certain types of cancer).

Herpes simplex virus vectors

HSV vectors are tropic for the central nervous system (CNS). Their major applications are expected to be in delivering genes into neurons for the treatment of neurological diseases, such as Parkinson's disease, and for treating CNS tumors.

Lentiviruses

The lentivirus family, which includes HIV (human immunodeficiency virus), are complex retroviruses that infect

macrophages and lymphocytes. Because of their ability to infect nondividing cells and to integrate into host cell chromosomes, considerable efforts are now being devoted to making lentivirus vectors for gene therapy.

Non-viral vector

Increasingly, concern has been expressed regarding the safety of viral vector systems. Therefore, attention has been focused toward studying alternative methods of gene transfer.

Liposomes

Liposomes are spherical vesicles composed of synthetic lipid bilayers which mimic the structure of biological membranes. The DNA to be transferred is packaged in vitro with the liposomes and used directly for transferring the DNA to a suitable target tissue in vivo.

Direct injection

In some cases, DNA can be injected directly with a syringe and needle into a specific tissue, such as muscle. This approach has been considered, for example, in the case of DMD, where early studies investigated intramuscular injection of a dystrophin minigene into a mouse model, for several months.

Receptor-mediated endocytosis

The DNA is coupled to a targeting molecule that can bind to a specific cell surface receptor, inducing endocytosis and transfer of the DNA into cells.

Inhibition of gene expression

One way of treating certain human disorders is to selectively inhibit the expression of a predetermined gene in vivo. In

principle, this general approach is particularly suited to treating cancers and infectious diseases, and some immunological disorders.

The expression of a selected gene might be inhibited by a variety of different strategies. One possible type of approach involves specific in vivo mutagenesis of that gene, altering it to a form that is no longer functional. Various techniques for selectively inhibiting expression of a specific gene have been devised, and include examples where expression is inhibited at all three major levels.

- Targeted inhibition of expression at the DNA level.
- Targeted inhibition of expression at the RNA level.
- Targeted inhibition of expression at the protein level.

Gene therapy for inherited disorders

Over the last two decades molecular genetic technologies have been spectacularly successful in identifying and characterizing novel disease genes, and in devising novel diagnostic tests for inherited disorders. In contrast, the dream of successfully applying molecular genetic technologies on a large scale to curing, or even treating disease has remained unfulfilled

However, some genetic disorders may not be so easy to treat as others. Certain single gene disorders will be more amenable to gene therapy approaches than others.

The first gene therapy trial for an inherited disease was initiated in 1990

The first gene therapy trial for an inherited disorder was initiated on 14 September 1990. The patient, Ashanthi DeSilva, was just 4

years old and was suffering from a very rare recessively inherited disorder, adenosine deaminase (ADA) deficiency. ADA is involved in the purine salvage pathway of nucleic acid degradation, and is a housekeeping enzyme which is synthesized in many different types of cell. An inherited deficiency of this enzyme has severe consequences in the case of T lymphocytes, one of the major classes of immune system cells. As a result, ADA^- patients suffer from severe combined immunodeficiency.

This severe disorder was particularly amenable to gene therapy for a variety of reasons: (1) the ADA gene is small, (2) had previously been cloned, (3) the target cells are T cells which are easily accessible and easy to culture, enabling ex vivo gene therapy and (4) the disorder is recessively inherited.

The novel ADA gene therapy approach involved essentially four steps:(i) cloning a normal ADA gene into a retroviral vector; (ii) transfecting the ADA recombinant into cultured ADA^- T lymphocytes from the patient; (iii) identifying the resulting ADA^+ T cells and expanding them in culture and (iv) re-implanting these cells in the patient.

This approach was designed to be a form of treatment which would need to be repeated on many occasions. Successful treatment would require high efficiency gene transfer into bone marrow stem cells and high levels of expression.

- Familial hypercholesterolemia (FH)
- Cystic fibrosis
- Duchenne muscular dystrophy
- Cancer gene therapies

Many different approaches can be used for cancer gene therapy In a few cases, the gene therapy approach has focused on targeting single genes, such as TP53 gene augmentation therapy and delivery of antisense KRAS genes in the case of some forms of non-small-cell lung cancer. In most cases, however, targeted killing of cancer cells has been conducted without knowing the molecular etiology of the cancer. Thus far, some significant advances have been made against local and metastatic tumor growth, but effective therapy awaits development of more effective methods to transfer and express transgenes or to induce antitumor responses .

The ethics of human gene therapy

Somatic gene therapy

All current gene therapy trials involve treatment for somatic tissues (somatic gene therapy). Somatic gene therapy, in principle, has not raised many ethical concerns. Many, therefore, view the ethics of somatic gene therapy to be at least as acceptable as, organ transplantation, and feel that ethical approval is appropriate for carefully assessed proposals.

Germline gene therapy

Germline gene therapy involving the genetic modification of germline cells (e.g. in the early zygote), is considered to be entirely different. It has been successfully practised on animals (e.g. to correct β-thalassemia in mice). However, thus far, it has not been sanctioned for the treatment of human disorders, and approval is unlikely to be given in the near future.

The lack of enthusiasm for the practice of germline gene therapy can be ascribed to three major reasons.

The imperfect technology

Germline gene therapy requires modification of the genetic material of chromosomes, and the used vector systems do not allow accurate control. However, in germline gene therapy, genetic modification has implications not just for a single cell: accidental insertion of an introduced gene or DNA fragment could result in a novel inherited pathogenic mutation.

The questionable ethics of germline modification

Genetic modification of human germline cells may have consequences not just for the individual whose cells were originally altered, but also for all individuals who inherit the genetic modification in subsequent generations. Germline gene therapy would inevitably mean denial of the rights of these individuals to any choice about whether their genetic constitution should have been modified in the first place .

The questionable need for germline gene therapy

A 100% chance of inheriting a harmful mutation could most likely occur in two ways. One is when an affected woman is **homoplasmic** for a harmful mutation in the mitochondrial genome and wishes to have a child. A second situation concerns inheritance of mutations in the nuclear genome. To have a 100% risk of inheriting a harmful mutation would require mating between a man and a woman both of whom have the same recessively inherited disease, an extremely rare occurrence.

Instead, the vast majority of mutations in the nuclear genome are inherited with at most a 50% risk (for dominantly inherited disorders) or a 25% risk (for recessively inherited disorders).

Since in vitro fertilization generally involves the production of several fertilized eggs, it would be much simpler to screen for normal eggs and select these for implantation, rather than to attempt genetic modification of fertilized eggs identified as carrying the harmful mutation.

Summary

The term gene therapy describes any procedure intended to treat a disease by genetically modifying the cells of a patient. Including infectious diseases inherited disorders and immune system disorders. Current gene therapy is exclusively **somatic** gene therapy, the introduction of genes into somatic cells of an affected individual. The prospect of human germline gene therapy raises a number of ethical concerns, and is currently not sanctioned.

Non-classical gene therapy: The idea to inhibit the expression of genes associated with the pathogenesis, or to correct a genetic defect and so restore normal gene expression.

1-Genetically engineered hormones

Recombinant human insulin and other cloned human gene products of medical interest have been produced commercially.

2-Genetically engineered antibody

Artificially produced therapeutic antibodies are designed to be *monospecific* and can recognize specific disease-associated antigens, leading to killing of the disease cells.

3-Genetically engineered vaccines

Recombinant DNA technology is also being applied to the construction of novel vaccines.

Classical gene therapy. to deliver genes to appropriate target cells with the aim of obtaining optimal expression of the introduced genes.

1. produce a product that the patient lacks.
2. kill diseased cells directly.
3. activate cells of the immune system so as to aid killing of diseased cells.

Two major general approaches are used in the transfer of genes for gene therapy: transfer of genes into patient cells outside of the body (*ex vivo*) or inside the body (*in vivo*).

Principles of gene transfer

Minigene: An artificial minigene (little form of the gene) may be used

Genes integrated into chromosomes

As progeny cells contain the introduced genes, amplifying them in cell culture and and transferring the cells back into the patient.

Nonintegrated genes
Some gene transfer systems are designed to insert genes into cells where they remain as extrachromosomal elements and may be expressed at high levels. For example, cancer gene therapies.

Vectors
Are the vehicle which carry the gene into the cells.

A- Viral: (1) Oncoretroviral (2)Adenovirus (3)Herpes simplex virus (4) Lentiviruses

B- Non-viral vector (1) Liposomes (2) Direct injection (3)Receptor-mediated endocytosis.

Inhibition of gene expression
One way of treating certain human disorders is to selectively inhibit the expression of a predetermined gene *in vivo.*

- Targeted inhibition of expression at the DNA level.
- Targeted inhibition of expression at the RNA level.
- Targeted inhibition of expression at the protein level.

Gene therapy for inherited disorders
Certain single gene disorders will be more amenable to gene therapy approaches than others.

The questionable ethics of germline modification
Genetic modification of human germline cells may have consequences for individuals who inherit the genetic modification in subsequent generations.

The questionable need for germline gene therapy
Since *in vitro* fertilization generally involves the production of several fertilized eggs, it would be much simpler to screen for normal eggs and select these for implantation, rather than to attempt genetic modification of fertilized eggs identified as carrying the harmful mutation.

GLOSSARY

Allele -- an alternative form of a gene; any one of several mutational forms of a gene. One of the variant forms of a gene at a particular locus, or location, on a chromosome. Different alleles produce variation in inherited characteristics such as hair color or blood type. In an individual, one form of the allele (the dominant one) may be expressed more than another form (the recessive one).

Alu repetitive sequence -- the most common dispersed repeated DNA sequence in the human genome accounting for 5% of human DNA. The name is derived from the fact that these sequences are cleaved by the restriction endonuclease Alu.

Amino acid sequence -- the linear order of the amino acids in a protein or peptide. A group of 20 different kinds of small molecules that link together in long chains to form proteins. Often referred to as the "building blocks" of proteins.

Amplification -- any process by which specific DNA sequences are replicated disproportionately greater than their representation in the parent molecules.

Aneuploidy -- state of having variant chromosome number (too many or too few). (i.e. Down syndrome, Turner syndrome).

Antisense--the non-coding strand in double-stranded DNA. The antisense strand serves as the template for mRNA synthesis.

Autosome -- a nuclear chromosome other than the X- and Y-chromosomes.

Autosomal dominant --a pattern of Mendelian inheritance whereby an affected individual possesses one copy of a mutant allele and one normal allele. (In contrast, recessive diseases require that the individual have two copies of the mutant allele.) Individuals with autosomal dominant diseases have a 50-50 chance of passing the mutant allele and hence the disorder onto their children. Examples of autosomal dominant diseases include Huntington's disease, neurofibromatosis, and polycystic kidney disease.

Barr body -- the condensed single X-chromosome seen in the nuclei of somatic cells of female mammals.

Base pair-- a pair of hydrogen-bonded nitrogenous bases (one purine and one pyrimidine) that join the component strands of the DNA double helix. Two bases which form a "rung of the DNA ladder." A DNA nucleotide is made of a molecule of sugar, a molecule of phosphoric acid, and a molecule called a base. The bases are the "letters" that spell out the genetic code. In DNA, the code letters are A, T, G, and C, which stand for the chemicals adenine, thymine, guanine, and cytosine, respectively. In base pairing, adenine always pairs with thymine, and guanine always pairs with cytosine.

BRCA1/BRCA2--the first breast cancer genes to be identified. Mutated forms of these genes are believed to be responsible for about half the cases of inherited breast cancer, especially

those that occur in younger women. Both are tumor suppressor genes.

Base sequence -- a partnership of organic bases found in DNA and RNA; adenine forms a base pair with thymine (or uracil) and guanine with cytosine in a double-stranded nucleic acid molecule.

Cancer-- diseases in which abnormal cells divide and grow unchecked. Cancer can spread from its original site to other parts of the body and can also be fatal if not treated adequately.

Candidate gene-- a gene, located in a chromosome region suspected of being involved in a disease, whose protein product suggests that it could be the disease gene in question.

Carrier -- an individual heterozygous for a single recessive gene.An individual who possesses one copy of a mutant allele that causes disease only when two copies are present. Although carriers not affected by the disease, two carriers can produce a child who has the disease.

cDNA -- complementary DNA produced from a RNA template by the action of RNA- dependent DNA polymerase.

cDNA library-- collection of DNA sequences generated from mRNA sequences. This type of library contains only protein-coding DNA (genes) and does not include any non-coding DNA.

CM (Centimorgan)-- a measure of genetic distance that tells how far apart two genes are. Generally one centimorgan equals about 1 million base pairs.

Centromere -- a region of a chromosome to which spindle traction fibers attach during mitosis and meiosis; the position of the centromere determines whether the chromosome is considered an acrocentric, metacentric or telomeric chromosome.

Chromosome -- in the eukaryotic nucleus, one of the threadlike structures consisting of chromatin and carry genetic information arranged in a linear sequence.

Chromosome banding -- a technique for staining chromosomes so that bands appear in a unique pattern particular to the chromosome.

Clone -- genetically engineered replicas of DNA sequences.

Cloned DNA -- any DNA fragment that passively replicates in the host organism after it has been joined to a cloning vector.

Cloning--the process of making copies of a specific piece of DNA, usually a gene. When geneticists speak of cloning, they do not mean the process of making genetically identical copies of an entire organism.

Codon -- a sequence of three nucleotides in mRNA that specifies an amino acid.

Contig-- a chromosome map showing the locations of those regions of a chromosome where contiguous DNA segments overlap. Contig maps are important because they provide the ability to study a complete, and often large segment of the genome by examining a series of overlapping clones which then provide an unbroken succession of information about that region.

Consanguinity -- genetic relationship. Consanguineous individuals have at least one common ancestor in the preceding few generations.

Conservative change -- an amino acid change that does not affect significantly the function of the protein.

Contiguous genes -- genes physically close on a chromosome that when acting together express a phenotype.

Cosmids -- plasmid vectors designed for cloning large fragments of eukaryotic DNA; the vector is a plasmid into which phage lambda cohesive end sites have been inserted.

Cytogenetics -- the study of chromosomes.

Cytogenetics map--the visual appearance of a chromosome when stained and examined under a microscope. Particularly important are visually distinct regions, called light and dark bands, which give each of the chromosomes a unique appearance. This feature allows a person's chromosomes to be studied in a clinical test known as a karyotype, which allows scientists to look for chromosomal alterations.

Cytosine--one of the four bases in DNA that make up the letters ATGC, cytosine is the "C". The others are adenine, guanine, and thymine. Cytosine always pairs with guanine.

Diploid--the number of chromosomes in most cells except the gametes. In humans, the diploid number is 46.

Deletion -- the loss of a segment of the genetic material from a chromosome.

DNA fingerprint technique -- a method employed to determine differences in amino acid sequences between related

proteins; relies upon the presence of a simple tandem-repetitive sequences that are scattered throughout the human genome.

DNA hybridization -- a technique for selectively binding specific segments of single-stranded (ss) DNA or RNA by base pairing to complementary sequences on ssDNA molecules that are trapped on a nitrocellulose filter.

DNA probe -- any biochemical used to identify or isolate a gene, a gene product, or a protein.

DNA Replication--the process by which the DNA double helix unwinds and makes an exact copy of itself.

DNA sequencing -- "plus and minus" or "primed synthesis" method, developed by Sanger, DNA is synthesized in vitro in such a way that it is radioactively labeled and the reaction terminates specifically at the position corresponding to a given base; the "chemical" method, ssDNA is subjected to several chemical cleavage protocols that selectively make breaks on one side of a particular base.

DOE -- Department of Energy.

Dominant -- alleles that determine the phenotype displayed in a heterozygote with another (recessive) allele.

Down syndrome -- a type of mental deficiency due to trisomy (three copies) of autosome 21, a translocation of 21 or mosaicism.

Double helix--the structural arrangement of DNA, which looks something like an immensely long ladder twisted into a helix,

or coil. The sides of the "ladder" are formed by a backbone of sugar and phosphate molecules, and the "rungs" consist of nucleotide bases joined weakly in the middle by hydrogen bonds.

Doublication--a particular kind of mutation: production of one or more copies of any piece of DNA, including a gene or even an entire chromosome

Duchenne/Becker muscular dystrophy -- the most common and severe form of muscular dystrophy; transmitted as an X-linked trait. X-linked recessive. Symptoms include onset at 2-5 years with difficulty with gait and stairs, enlarged calf muscles, progression to wheelchair by adolescence, shortened life span.

Dystonia -- neurologic condition involving repeated twisting and movement. Involves a variety of muscle groups. Intelligence not effected. Three forms: childhood - autosomal dominant, autosomal recessive, adult-acquired.

Dwarfism -- conditions of short stature with adult height under 4'10" as adult, usually with normal intelligence and lifespan.

Ehlers Danlos Syndrome--- connective tissue condition including problems with tendons, ligaments, skin, bones, cartilage, and membranes surrounding blood vessels and nerves. Symptoms include joint laxity, elastic skin, dislocations. Many forms: autosomal dominant, autosomal recessive, X-linked forms.

Electrophoresis--The process in which molecules (such as proteins, DNA, or RNA fragments) can be separated according to size and electrical charge by applying an electric

current to them. The current forces the molecules through pores in a thin layer of gel, a firm jelly-like substance. The gel can be made so that its pores are just the right dimensions for separating molecules within a specific range of sizes and shapes. Smaller fragments usually travel further than large ones. The process is sometimes called gel electrophoresis.

Endonuclease -- an enzyme that breaks the internal phosphodiester bonds in a DNA molecule.

Ethics -- the study of fundamental principles which defines values and determines moral duty and obligation.

Euchromatin -- the chromatin that shows the staining behavior characteristic of the majority of the chromosomal complement.

Eugenics -- the improvement of humanity by altering its genetic composition by encouraging breeding of those presumed to have desirable genes.

Exons -- portion of a gene included in the transcript of a gene and survives processing of the RNA in the cell nucleus to become part of a spliced messenger of a structural RNA in the cell cytoplasm; an exon specifies the amino acid sequence of a portion of the complete polypeptide.

FISH -- florescent in situ hybridization: a technique for uniquely identifying whole chromosomes or parts of chromosomes using florescent tagged DNA. A process which vividly paints chromosomes or portions of chromosomes with fluorescent molecules. This technique is useful for identifying chromosomal abnormalities and gene mapping.

5' - end -- the end of a polynucleotide with a free (or phosphorylated or capped) 5' - hydroxyl group; transcription/translation begins at this end.

Fragile sites -- a non-staining gap of variable width that usually involves both chromatids and is always at exactly the same point on a specific chromosome derived from an individual or kindred.

Fragile-X syndrome -- X-linked trait; the second most common identifiable cause of genetic mental deficiency.

Gamete -- an haploid cell.gel electrophoresis the process by which nucleic acids (DNA or RNA) or proteins are separated by size according to movement of the charged molecules in an electrical field.

Gene -- a hereditary unit that occupies a certain position on a chromosome; a unit that has one or more specific effects on the phenotype, and can mutate to various allelic forms.

Gene amplification -- any process by which specific DNA sequences are replicated disproportionately greater than their representation in the parent molecules; during development, some genes become amplified in specific tissues.

Gene map -- the linear arrangement of mutable sites on a chromosome as deduced from genetic recombination experiments.

Gene therapy -- addition of a functional gene or group of genes to a cell by gene insertion to correct an hereditary disease.

Gene Transfer--Insertion of unrelated DNA into the cells of an organism. There are many different reasons for gene transfer:

for example, attempting to treat disease by supplying patients with therapeutic genes. There are also many possible ways to transfer genes. Most involve the use of a vector, such as a specially modified virus that can take the gene along when it enters the cell.

Genetic code(ATGC)--The instructions in a gene that tell the cell how to make a specific protein. A, T, G, and C are the "letters" of the DNA code; they stand for the chemicals adenine, thymine, guanine, and cytosine, respectively, that make up the nucleotide bases of DNA. Each gene's code combines the four chemicals in various ways to spell out 3-letter "words" that specify which amino acid is needed at every step in making a protein

Genetic counseling -- the educational process that helps individuals, couples, or families to understand genetic information and issues that may have an impact on them.

Genetic linkage map -- a chromosome map showing the relative positions of the known genes on the chromosomes of a given species.

Genetic screening -- testing groups of individuals to identify defective genes capable of causing hereditary conditions.

Genetic marker--A segment of DNA with an identifiable physical location on a chromosome and whose inheritance can be followed. A marker can be a gene, or it can be some section of DNA with no known function. Because DNA segments that lie near each other on a chromosome tend to be inherited together, markers are often used as indirect ways of

tracking the inheritance pattern of a gene that has not yet been identified, but whose approximate location is known.

Genetic variation -- a phenotypic variance of a trait in a population attributed to genetic heterogeneity.

Genome -- all of the genes carried by a single gamete; the DNA content of an individual, which includes all 44 autosomes, 2 sex chromosomes, and the mitochondrial DNA.

Genotype -- genetic constitution of an organism.

Germ cell -- a sex cell or gamete (egg or spermatozoan).Haldane equation

Haldane's law -- the generalization that if first generation hybrids are produced between two species, but one sex is absent, rare, or sterile, that sex is the heterogamic sex.

Heterozygote -- having two alleles that are different for a given gene.

Heterogeneity -- the production of identical or similar phenotypes by different genetic mechanisms.

HGP -- Human Genome Project.

High conservative sequance--A DNA sequence that is very similar in several different kinds of organisms. Scientists regard these cross species similarities as evidence that a specific gene performs some basic function essential to many forms of life and that evolution has therefore conserved its structure by permitting few mutations to accumulate in it.

Homologous chromosomes -- chromosomes that pair during meiosis; each homologue is a duplicate of one chromosome from each parent.

Homologous recombination-- The exchange of pieces of DNA during the formation of eggs and sperm. Recombination allows the chromosomes to shuffle their genetic material, increasing the potential of genetic diversity. Homologous recombination is also known as crossing over.

Human artificial chromosome-- A vector used to transfer or express large fragments of human DNA. HACs behave and are constructed like human chromosomes.

Homozygote -- having identical alleles at one or more loci in homologous chromosome segments.

Housekeeping genes -- those genes expressed in all cells because they provide functions needed for sustenance of all cell types.

HUGO -- Human Genome Organization.

Huntington disease -- a disease characterized by irregular, spasmodic involuntary movements of the limbs and facial muscles, mental deterioration and death, usually within 20 years of the onset of symptoms.

Hybridization -- the pairing of a single-stranded, labeled probe (usually DNA) to its complementary sequence.

Imprinting -- a chemical modification of a gene allele which can be used to identify maternal or paternal origin of chromosome.

Incomplete penetrance -- the gene for a condition is present, but not obviously expressed in all individuals in a family with the gene.

In situ hybridization -- hybridization of a labeled probe to its complementary sequence within intact, banded chromosomes. The base pairing of a sequence of DNA to metaphase chromosomes on a microscope slide.

Insertion-- A type of chromosomal abnormality in which a DNA sequence is inserted into a gene, disrupting the normal structure and function of that gene.

Introns -- a segment of DNA (between exons) that is transcribed into nuclear RNA, but are removed in the subsequent processing into mRNA.

Isochromosome -- a metacentric chromosome produced during mitosis or meiosis when the centromere splits transversely instead of longitudinally; the arms of such chromosome are equal in length and genetically identical, however, the loci are positioned in reverse sequence in the two arms.

Linkage-- The association of genes and/or markers that lie near each other on a chromosome. Linked genes and markers tend to be inherited together.

Klinefelter syndrome -- an endocrine condition caused by a an extra X-chromosome (47,XXY); characterized by the lack of normal sexual development and testosterone, leading to infertility and adjustment problems if not detected and treated early.

Karyotype -- a set of photographed, banded chromosomes arranged in order from largest to smallest. The chromosomal complement of an individual, including the number of chromosomes and any abnormalities. The term is also used to refer to a photograph of an individual's chromosomes.

Lligase -- an enzyme that functions in DNA repair.

Linkage -- analysis of pedigree the tracking of a gene through a family by following the inheritance of a (closely associated) gene or trait and a DNA marker.

Locus--The place on a chromosome where a specific gene is located, a kind of address for the gene. The plural is "loci," not "locuses."

Mandelian inheretance--Manner in which genes and traits are passed from parents to children. Examples of Mendelian inheritance include autosomal dominant, autosomal recessive, and sex-linked genes.

Marfan syndrome -- autosomal dominant condition of connective tissue; affects the skeletal, ocular and cardiovascular systems.

Marker -- a gene with a known location on a chromosome and a clear-cut phenotype, used as a point of reference when mapping a new mutant.

Meiosis -- the doubling of gametic chromosome number.

Messenger RNA(mRNA) -- Template for protein synthesis. Each set of three bases, called codons, specifies a certain protein in the sequence of amino acids that comprise the protein. The sequence of a strand of mRNA is based on the sequence of a complementary strand of DNA.

Methylation -- addition of a methyl group (-CH3) to DNA or RNA.

Microarray technology--A new way of studying how large numbers of genes interact with each other and how a cell's regulatory networks control vast batteries of genes simultaneously. The method uses a robot to precisely apply tiny droplets containing functional DNA to glass slides. Researchers then attach fluorescent labels to DNA from the cell they are studying. The labeled probes are allowed to bind to complementary DNA strands on the slides. The slides are put into a scanning microscope that can measure the brightness of each fluorescent dot; brightness reveals how much of a specific DNA fragment is present, an indicator of how active it is.

Microsatellite--Repetitive stretches of short sequences of DNA used as genetic markers to track inheritance in families

Missense mutation -- a change in the base sequence of a gene that alters or eliminates a protein.

Mitochondrial DNA -- the mitochondrial genome consists of a circular DNA duplex, with 5 to 10 copies per organelle.

Mitosis -- nuclear division.

mRNA -- messenger RNA; an RNA molecular that functions during translation to specify the sequence of amino acids in a nascent polypeptide.

Multifactorial -- a characteristic influenced in its expression by many factors, both genetic and environmental.

Mutation -- process by which genes undergo a structural change

Non coding DNA--The strand of DNA that does not carry the information necessary to make a protein. The non-coding

strand is the mirror image of the coding strand and is also known as the antisense strand.

Nonsense mutation -- a mutation in which a codon is changed to a stop codon, resulting in a truncated protein product.

Northern analysis -- a technique for transferring electrophoretically resolved RNA segments from an agarose gel to a nitrocellulose filter paper sheet via capillary action.

Nucleotide -- one of the monomeric units from which DNA or RNA polymers are constructed; consists of a purine or pyrimidine base, a pentose sugar and a phosphoric acid group.

Oncogenes -- genes involved in cell cycle control (growth factors, growth factor regulator genes, etc), a mutation can lead to tumor growth

Oligonucleotide --oligonucleotide, short sequence of single-stranded DNA or RNA. Oligos are often used as probes for detecting complementary DNA or RNA because they bind readily to their complements.

Oncogene--a gene that is capable of causing the transformation of normal cells into cancer cells.

Oncovirus--A class of retroviruses that cause a cell to become cancerous.

Osteogenesis imperfecta -- a condition also known as brittle bone disease; characterized by a triangular shaped face with yellowish brown teeth, short stature and stunted growth,

scoliosis, high pitched voice, excessive sweating and loose joints.

Parthenogenesis -- the development of an individual from an egg without fertilization.

P53-- a gene which normally regulates the cell cycle and protects the cell from damage to its genome. Mutations in this gene cause cells to develop cancerous abnormalities.

PCR -- polymerase chain reaction; a technique for copying the complementary strands of a target DNA molecule simultaneously for a series of cycles until the desired amount is obtained.

Pedigree -- a diagram of the heredity of a particular trait through many generations of a family.

Phenotype -- observable characteristics of an organism produced by the organism's genotype interacting with the environment.

Physical map -- map where the distance between markers is the actual distance, such as the number of base pairs.

PKU -- phenylketonuria, an enzyme deficiency condition characterized by the inability to convert one amino acid, phenylalanine, to another, tyrosine, resulting in mental deficiency. plasmid double-stranded, circular, bacterial DNA into which a fragment of DNA from another organism can be inserted.

Pleiotropy -- the phenomenon of variable phenotypes for a number of distinct and seemingly unrelated phenotypic effects.

Polycystic kidney disease (PKD) -- a group of conditions characterized by fluid filled sacs that slowly develop in both kidneys, eventually resulting in kidney malfunction.

Polymerase -- any enzyme that catalyzes the formation of DNA or RNA from deoxyribonucleotides or ribonucleotides.

Polymorphism--a common variation in the sequence of DNA among individuals

Prader-Willi syndrome -- a condition characterized by obesity and insatiable appetite, mental deficiency, small genitals, and short stature. May be deletion of #15 chromosome.

Positional cloning--a process which, through gene mapping techniques, is able to locate a gene responsible for a disease when little or no information is known about the biochemical basis of the disease.

Predisposition -- to have a tendency or inclination towards something in advance.

Presymptomatic diagnosis -- diagnosis of a genetic condition before the appearance of symptoms.

Primer -- nucleotides used in the polymerase chain reaction to initiate DNA synthesis at a particular location.

Probability -- the long term frequency of an event relative to all alternative events, and usually expressed as decimal fraction.

Proband -- individual in a family who brought the family to medical attention.

Probe -- single-stranded DNA labeled with radioactive isotopes or tagged in other ways for ease in identification.

Prognosis -- prediction of the course and probable outcome of a disease.

Psoudogene--a sequence of DNA that is very similar to a normal gene but that has been altered slightly so it is not expressed. Such genes were probably once functional but over time acquired one or more mutations that rendered them incapable of producing a protein product.

Recessive -- a gene that is phenotypically manifest in the homozygous state but is masked in the presence of a dominant allele.

Recombination -- the natural process of breaking and rejoining DNA strands to produce new combinations of genes and, thus, generate genetic variation. Gene crossover during meiosis.

Repeat sequences -- the length of a nucleotide sequence that is repeated in a tandem cluster.

Restriction enzymes--enzymes that recognize a specific sequence of double-stranded DNA and cut the DNA at that site. Restriction enzymes are often referred to as molecular scissors.

Restriction Fragment Length Polymorphism (RFLP)-- genetic variations at the site where a restriction enzyme cuts a piece of DNA. Such variations affect the size of the resulting fragments. These sequences can be used as markers on physical maps and linkage maps. RFLP is also pronounced "rif" lip.

Retinitis pigmentosa -- group of hereditary ocular disorders with progressive retinal degeneration. Autosomal dominant, autosomal recessive, and X-linked forms.

Retinoblastoma -- a childhood malignant cancer of the retina of the eye. reverse transcriptase viral enzyme used to make cDNA.

RFLP -- restriction fragment length polymorphism; variations occurring within a species in the length of DNA fragments generated by a species endonuclease.

Ribosomal protein -- one of the ribonucleoprotein particles that are the sites of translation.

Sanger sequence -- "plus and minus" or "primed synthesis" method; DNA is synthesized so it is radioactively labeled and the reaction terminates specifically at the position corresponding to a given base.

Selection -- the process of determining the relative share allotted individuals of different genotypes in the propagation of a population; the selective effect of a gene can be defined by the probability that carriers of the gene will reproduce.

Sequence tagged site--A short DNA segment that occurs only once in the human genome and whose exact location and order of bases are known. Because each is unique, STSs are helpful for chromosome placement of mapping and sequencing data from many different laboratories. STSs serve as landmarks on the physical map of the human genome

Sex determination -- the mechanism in a given species by which sex is determined; in many species sex is determined at fertilization by the nature of the sperm that fertilizes the egg.

Single Nucleotide Polymorphism (SNPs)--Common, but minute, variations that occur in human DNA at a frequency of one every 1,000 bases. These variations can be used to track inheritance in families. SNP is pronounced "snip".

Sickle cell anemia -- an hereditary, chronic form of hemolytic anemia characterized by breakdown of the red blood cells; red blood cells undergo a reversible alteration in shape when the oxygen tension of the plasma falls slightly and a sickle-like shape forms.

Somatic cell hybrid -- hybrid cell line derived from two different species; contains a complete chromosomal complement of one species and a partial chromosomal complement of the other; human/hamster hybrids grow and divide, losing human chromosomes with each generation until they finally stabilize, the hybrid cell line established is then utilized to detect the presence of genes on the remaining human chromosome.

Somatic mutation -- a mutation occurring in any cell that is not destined to become a germ cell; if the mutant cell continues to divide, the individual will come to contain a patch of tissue of genotype different from the cells of the rest of the body.

Southern blotting -- a technique for transferring electrophoretically resolved DNA segments from an agarose

gel to a nitrocellulose filter paper sheet via capillary action; the DNA segment of interest is probed with a radioactive, complementary nucleic acid, and its position is determined by autoradiography.

Spina bifida -- a congenital condition that results from altered fetal development of the spinal cord, part of the neural plate fails to join together and bone and muscle are unable to grow over this open section.

Syndrome -- a recognizable pattern or group of multiple signs, symptoms or malformations that characterize a particular condition; syndromes are thought to arise from a common origin and result from more than one developmental error during fetal growth.

Tay-Sachs disease -- a fatal degenerative disease of the nervous system due to a deficiency of hexosamidase A, causing mental deficiency, paralysis, mental deterioration, and blindness; found primarily but not exclusively among Ashkenazi Jews. Autosomal recessive.

Teratogens -- any agent that raises the incidence of congenital malformations.

3' - end -- the end of a polynucleotide with a free (or phosphorylated) 3' - hydroxyl group.

Trait -- any detectable phenotypic property of an organism.

Transduction -- the transfer of bacterial genetic material from one bacterium to another using a phage as a vector.

Transferase -- enzymes that catalyze the transfer of functional groups between donor and acceptor molecules.

Transcription -- the formation of an RNA molecule upon a DNA template by complementary base pairing.

Translation -- the formation of a polypeptide chain in the specific amino acid sequence directed by the genetic information carried by mRNA.

Translocation -- a chromosome aberration which results in a change in position of a chromosomal segment within the genome, but does not change the total number of genes present.

Triplet code -- a code in which a given amino acid is specified by a set of three nucleotides.

Tumor suppressor gene -- genes that normally function to restrain the growth of tumors; the best understood case is for hereditary retinoblastoma.

Transgenic organism -- one into which a cloned genetic material has been experimentally transferred, a subset of these foreign gene express themselves in their offspring.

Turner syndrome--- a chromosomal condition in females (usually 45,XO) due to monosomy of the X- chromosome; characterized by short stature, failure to develop secondary sex characteristics, and infertility.

VNTR -- variable number tandem repeats; any gene whose alleles contain different numbers of tandemly repeated oligonucleotide sequences.

Vector -- a self-replicating DNA molecule that transfers a DNA segment between host cells.

Western blotting analysis -- a technique used to identify a specific protein; the probe is a radioactively labeled antibody raised against the protein in question.

X-inactivation -- the repression of one of the two X-chromosomes in the somatic cells of females as a method of dosage compensation; at an early embryonic stage in the normal female, one of the two X-chromosomes undergoes inactivation, apparently at random, from this point on all descendent cells will have the same X-chromosome inactivated as the cell from which they arose, thus a female is a mosaic composed of two types of cells, one which expresses only the paternal X-chromosome, and another which expresses only the maternal X-chromosome.

XYY syndrome -- genetic condition in males with extra Y chromosome (in 1 in 1000 male births). Symptoms: tall stature (over 6'), may including sterility, developmental delay, learning problems.

YAC -- yeast artificial chromosome; a linear vector into which a large fragment of DNA can be inserted; the development of YAC's in 1987 has increased the number of nucleotides which can be cloned.

رقم الإيداع ١٨٦٤٣ / ٢٠٠٥